

**Transcriptome and secretome analysis of intra-mammalian life-stages of the emerging helminth pathogen, *Calicophoron daubneyi* reveals adaptation to a unique host environment.**

Kathryn M. Huson<sup>1</sup>, Erwan Atcheson<sup>1</sup>, Nicola A.M. Oliver<sup>1</sup>, Philip Best<sup>1</sup>, Jason P. Barley<sup>2</sup>, Robert E.B. Hanna<sup>2</sup>, Tom N. McNeilly<sup>3</sup>, Yongxiang Fang<sup>4</sup>, Sam Haldenby<sup>4</sup>, Steve Paterson<sup>4</sup>  
& Mark W. Robinson<sup>1</sup>

<sup>1</sup>School of Biological Sciences, Queen's University Belfast, 19 Chlorine Gardens, Belfast, Northern Ireland; <sup>2</sup>Veterinary Sciences Division, Agri-Food and Biosciences Institute, Belfast, Northern Ireland; <sup>3</sup>Disease Control Department, Moredun Research Institute, UK; <sup>4</sup>Centre for Genomic Research, University of Liverpool, UK

**Running Title:** The transcriptome and secretome of *C. daubneyi*

**Correspondence:** [mark.robinson@qub.ac.uk](mailto:mark.robinson@qub.ac.uk) (M.W. Robinson).

Tel: 02890 972120; Fax: 02890 975877

**Key words:** *Calicophoron daubneyi*, rumen fluke, paramphistome, secretome, transcriptome, diagnostic, coproantigen, ELISA.

## Abbreviations

NEJ - newly excysted juvenile

E/S - excretory/secretory

BLAST - Basic local alignment search tool

TPM - transcripts per million

FA - formic acid

emPAI – exponentially modified protein abundance index

$\alpha$ -CdE/S – anti-*C. daubneyi* excretory-secretory protein antibody

$\alpha$ -CdE/S-B - biotinylated anti-*C. daubneyi* excretory-secretory protein antibody

WWE, whole worm extract

AbD - antibody diluent

CRISP - cysteine-rich secretory protein

TCA - tricarboxylic acid cycle

HDM - helminth defence molecule

FaBP - fatty-acid binding protein

VAL - venom allergen-like

GST - glutathione *S*-transferase

LPS – lipopolysaccharide

PAMP - pathogen-associated molecular pattern

## Summary

Paramphistomosis, caused by the rumen fluke, *Calicophoron daubneyi*, is a parasitic infection of ruminant livestock which has seen a rapid rise in prevalence throughout Western Europe in recent years. Following ingestion of metacercariae (parasite cysts) by the mammalian host, newly-excysted juveniles (NEJs) emerge and invade the duodenal submucosa which causes significant pathology in heavy infections. The immature larvae then migrate upwards, along the gastrointestinal tract, and enter the rumen where they mature and begin to produce eggs. Despite their emergence, and sporadic outbreaks of acute disease, we know little about the molecular mechanisms used by *C. daubneyi* to establish infection, acquire nutrients and to avoid the host immune response. Here, transcriptome analysis of four intra-mammalian life-cycle stages, integrated with secretome analysis of the NEJ and adult parasites (responsible for acute and chronic disease respectively), revealed how the expression and secretion of selected families of virulence factors and immunomodulators are regulated in accordance with fluke development and migration. Our data show that whilst a family of cathepsins B with varying S2 sub-site residues (indicating distinct substrate specificities) are differentially secreted by NEJs and adult flukes, cathepsins L and F are secreted in low abundance by NEJs only. We found that *C. daubneyi* has an expanded family of aspartic peptidases, which is up-regulated in adult worms, although they are underrepresented in the secretome. The most abundant proteins in adult fluke secretions were helminth defence molecules (HDMs) that likely establish an immune environment permissive to fluke survival and/or neutralise pathogen-associated molecular patterns (PAMPs) such as bacterial lipopolysaccharide in the microbiome-rich rumen. The distinct collection of molecules secreted by *C. daubneyi* allowed the development of the first coproantigen-based ELISA for paramphistomosis which, importantly, did not recognise antigens from other helminths commonly found as co-infections with rumen fluke.

## Introduction

Infections by parasitic fluke are an important animal health and production concern for livestock producers worldwide. In the UK, and throughout Europe, the liver fluke (*Fasciola hepatica*) has historically been a major focus for producers. In recent years, however, there has been a sharp increase in both the incidence and severity of rumen fluke (the paramphistome *Calicophoron daubneyi*) infections in both sheep and cattle (1). Paramphistomes are endemic to sub-tropical and tropical regions and are believed to have been carried into Western Europe from North Africa via the movement of ruminant livestock (2). Although the exact reasons for the rise in rumen fluke infections in Europe are not fully understood, the increase in warm wet summers and mild winters - conditions that favour *Galba truncatula*, the snail intermediate host of *C. daubneyi* - are thought to be a major contributing factor (3).

Acute, clinical paramphistomosis is caused when grazing stock ingest large numbers of metacercariae from pasture, which then excyst *en masse* in the duodenum (4). The newly excysted juvenile (NEJ) flukes then migrate into the intestinal submucosa causing significant damage to the host tissue (5). Large areas of damaged small intestine may haemorrhage, causing significant blood loss and hypoalbuminaemia, frequently resulting in mortality at this point (6, 7). After a period spent feeding on the host tissue in the small intestine, immature paramphistomes migrate to the rumen where they mature, and infections become patent (8). Although chronic infections are generally seen as well tolerated, post-mortem observations have noted both rumenitis and abomasitis in infected animals (9, 10), along with atrophy of the rumen papillae (1, 10).

Despite a prevalence of 55 % - 77 % (1), clinical disease is still relatively rare in the UK/Ireland. However, fatal disease outbreaks, linked to significant immature parasite burdens have been reported in both sheep and cattle in recent years (4, 6, 10 - 13). Control of

fluke infection currently relies on anthelmintic drugs. Whilst several drugs show efficacy against liver fluke, only one anthelmintic (oxyclozanide) is effective against rumen fluke (14, 15). This makes correct diagnosis imperative. However, detection of rumen fluke infection currently requires either examination of animals at post-mortem or labour-intensive faecal egg counts which only detect chronic infection due to the presence of egg-producing adult flukes. Thus, the development of new tools for more rapid diagnosis of *C. daubneyi* (including acute infection) are required.

Compared to other helminths of veterinary importance, *C. daubneyi* remains a poorly studied species. We have much to learn about its basic biology and interactions with the ruminant host, particularly the infective juvenile stages (1). Here we have performed the first transcriptomic analysis of all four major intra-mammalian life-cycle stages of *C. daubneyi*. These data, coupled with secretome analysis of the infective and adult stages that are associated with acute and chronic disease respectively, have provided a greater understanding of the key biochemical and molecular mechanisms underpinning *C. daubneyi* infectivity, migration and development within its ruminant host. Our data reveal how the parasite regulates expression and secretion of a collection of molecules required for tissue invasion, nutrition and modulation of the host immune response according to fluke development and exposure to different host microenvironments (i.e. the duodenum and rumen). Owing to the unique composition of *C. daubneyi* secretions compared with those of *F. hepatica* (often found as co-infections), the diagnostic potential of these molecules was investigated. Accordingly, we present the first ELISA-based assay capable of detecting *C. daubneyi* antigens in faecal samples from naturally-infected cattle. Our data represent an important foundation for future studies aimed at understanding rumen fluke infectivity and the development of future treatment options and diagnostic tests.

## Materials and Methods

### *Experimental Design and Statistical Rationale*

For transcriptomics studies, 3 biological replicates of each life-cycle stage were prepared for Illumina RNAseq, with 1 replicate used for PacBio sequencing for use as full-length transcript scaffold data during assembly. Three biological replicates were also used for the proteomics study of excretory/secretory (E/S) products analysed by LC-MS/MS. ELISA data are representative of three independent experiments.

### *Source of parasite material and RNA extraction*

To obtain RNA samples, *C. daubneyi* adults and newly migrated parasites were collected from the rumen of a naturally-infected cow in a local abattoir (Dungannon, Northern Ireland). Individual adults and small groups (approx. 6) of newly migrated flukes were rinsed briefly in sterile warm (39°C) PBS and immediately placed into TRIzol reagent (Life Technologies) to preserve mRNA. Immature parasites were removed from the duodenum of a recently deceased bovine host during post mortem examination and approximately 200 mg of parasite material placed directly into TRIzol reagent. Rumen fluke metacercariae (Ridgeway Research) were excysted *in vitro* and the resulting NEJ parasites cultured in RPMI-1640 medium for 24 hours as previously described (16). The culture media was removed and TRIzol reagent was added directly to groups of 500 NEJs. All samples were stored in TRIzol at -80°C until RNA extraction. Samples were homogenised in TRIzol and total RNA was extracted following the manufacturer's protocol. RNA yields were quantified using a NanoDrop™ spectrophotometer and shipped to the Centre for Genomic Research (University of Liverpool, UK) for library construction. Prior to sequencing, the rumen fluke species was confirmed by PCR using *C. daubneyi*-specific primers targeting a 885 bp region of the

cytochrome oxidase 1 gene; Cd\_Cox1F 5'-TGGAGAGTTTGGCGTCTTTT-3' and Cd\_Cox1R 5'-CCATCTTCCACCTCATCTGG-3' as previously described (17).

#### *PacBio Isoseq library preparation and sequencing*

Full length cDNA was made from 1 µg of total RNA from each sample using the Teloprime full length cDNA amplification kit from Lexogen. The final second strand product was amplified for 20 cycles and converted to a PacBio SMRTbell library using the SMRTbell Template Prep Kit 1.0. Each sample was run on a single Sequel SMRT cell v2 using Magbead loading and 10 hour movie times. PacBio reads were analysed using the Iso-seq pipeline (SMRTlink version 4.0.0.190159), with default parameters. Each sample was processed independently. Subsequently, high and low quality full-length isoforms from each sample were clustered using CD-HIT (18). Cluster representatives were polished with Arrow (19) and reclustered with cd-hit-est, as above but with more relaxed parameters to limit over-clustering (-aL 0.97). These two steps were iteratively performed to generate a set of 27,949 isoforms, which were subsequently filtered to remove any sequences that (a) had a mean base quality score of less than 50 or (b) had more than 1% of bases with a quality score below 10. This yielded a final set of 27,254 isoforms.

#### *Annotation of transcripts*

Putative proteins sequences were obtained for isoforms using TransDecoder (version 3.0.1; (20)), selecting for a minimum amino acid sequence length of 50. Sequences were annotated with input for the Trinotate pipeline (version 3.1.0; (20)), according to the proposed guidelines from the authors. The basic local alignment search tool (BLAST) (21) was used to align nucleotide and protein sequences against UniProt (release 2017\_11), TrEMBL and

against proteins encoded by *F. hepatica* gene models available at WormbaseParasite ([ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS2/species/fasciola\\_hepatica/PRJEB6687/fasciola\\_hepatica.PRJEB6687.WBPS2.protein.fa.gz](ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS2/species/fasciola_hepatica/PRJEB6687/fasciola_hepatica.PRJEB6687.WBPS2.protein.fa.gz)). All BLAST alignments which yielded e-values of  $1 \times 10^{-5}$  or lower were reported. Predicted amino acid sequences were compared to Pfam-A (release 31.0) (22), using hmmscan (version 3.1b1; (23)). Signal peptides and transmembrane domains were predicted using SignalP (version 4.1; (24)) and tmhmm (version 2.0; (25)).

#### *Illumina RNASeq library preparation and sequencing*

Triplicate total RNA samples (from *C. daubneyi* NEJs, immature intestinal flukes, newly-migrated flukes and mature adults) were treated with Turbo DNase (Life Technologies) to remove all contaminating DNA. 1 µg total RNA was selected for poly A using NEBNext® Poly(A) mRNA Magnetic Isolation Module. RNA-Seq libraries were prepared from the enriched material using the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® using 12 cycles of amplification. Final libraries were pooled in equimolar amounts using the Qubit and Bioanalyzer data and checked by quantitative PCR for template loading calculations. DNA was diluted to a loading concentration of 300 pM. The RNA libraries were sequenced on an Illumina® HiSeq 4000 platform with version 1 chemistry using sequencing by synthesis (SBS) technology to generate 2 x 150 bp paired-end reads.

#### *RNAseq data processing and quality filtering*

Base calling and de-multiplexing of indexed reads was performed by CASAVA version 1.8.2 (Illumina). Sequences were trimmed using Cutadapt version 1.2.1 (26), to remove any adaptor sequences matching with 3 or more bases. The reads were further trimmed to remove



low quality bases, using Sickle version 1.200 with a minimum window quality score of 20. Illumina reads were aligned to the final Isoseq isoforms using Bowtie2 version 2.1.0 (27), accepting all global concordantly paired alignments.

### *Differential expression analysis*

Read alignment data was used as input to BitSeq (28), to estimate raw expression values for each transcript in the dataset. Differential transcript expression analysis was applied to the read count data which was conducted in the R environment, using the DESeq2 package (29). Transcripts per million (TPM) values were calculated by dividing the counts to each transcript by the length of the transcript and then normalised further to generate counts per million reads. Transcripts were clustered according to their log-fold changes in expression between sample groups, using k-means clustering functions in R.

### *Discovery and characterisation of C. daubneyi gene families*

*C. daubneyi* gene families and specific homologs of interest were identified using BlastP (1.0 x 10<sup>-6</sup> e-value cut-off) run locally with BioEdit (30) using published query sequences from other trematode species (*F. hepatica*, *Fasciola gigantica*, *Opisthorchis viverrini*, *Clonorchis sinensis* and *Schistosoma spp.*). Duplicate and truncated sequences were manually removed from the BLAST hits and transcript TPM values were extracted for the remaining transcripts at each of the 4 lifecycle stages to allow differential expression analysis across the intra-mammalian stages of the *C. daubneyi* lifecycle and used to produce heatmaps via the heatmapper.ca expression tool (31). Helical wheel analysis was conducted on *C. daubneyi* HDM family members ([http://tcdb.org/progs/helical\\_wheel.php](http://tcdb.org/progs/helical_wheel.php)) to identify amphipathic regions.

Neighbour-Joining phylogenetic trees were constructed for selected protein sequences using MEGA6 (32). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed (33). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Poisson correction method (34) and are in the units of the number of amino acid substitutions per site.

#### *Collection of parasite excretory/secretory (E/S) products*

To prepare secretions, adult rumen flukes were thoroughly washed (3 x 5 minutes) with warm (39°C) sterile PBS to void their gut contents and to remove any host contaminants. The integrity and movement of the flukes was monitored by microscopy and any found to be damaged or dead were discarded prior to culture. Flukes were then maintained in RPMI-1640 culture medium containing 0.1% glucose, 100 U penicillin and 100 µg/ml streptomycin (Sigma-Aldrich), at 1 worm/ml for 5 h at 39°C. Whole E/S was also collected from *C. daubneyi* NEJs (in groups of 500 parasites) as described previously (16). Excysted NEJs were cultured in 1 mL RPMI-1640 supplemented with 100 IU mL<sup>-1</sup> penicillin and 100 mg mL<sup>-1</sup> streptomycin for 24 h at 39°C. The culture media was recovered and centrifuged for 30 minutes at 1500 x *g* to remove large debris. Adult and NEJ E/S fractions were concentrated (approx. 10-fold) using Amicon Ultra Centrifugal Filter Units with a 3 kDa molecular weight cut-off (Millipore, UK) and stored at – 80°C until further analysis.

E/S proteins were reduced with 2 mM DTT in 50 mM NH<sub>4</sub>HCO<sub>3</sub> at 60°C for 20 min and alkylated with 5 mM iodoacetamide at room temperature (18-21°C) in the dark for 30

min. E/S samples were incubated with 100 ng/ $\mu$ L sequencing grade trypsin (Promega) overnight at 37°C. The digestions were stopped by the addition of TFA to a final concentration of 0.1%.

#### *Mass spectrometry analysis of C. daubneyi proteins*

E/S proteins were analysed in biological triplicate. Tryptic peptides were dried in a vacuum centrifuge and reconstituted with 10  $\mu$ L of 0.1% TFA before analysis by LC-MS/MS. Peptides is 5  $\mu$ L of the resulting suspension were purified using an Acclaim PepMap 100 column (C18, 100  $\mu$ M x 2 cm) prior to delivery to an analytical column (Eksigen C18-CL NanoLC Column, 3  $\mu$ m; 75  $\mu$ m x 15 cm) equilibrated in 5% acetonitrile/0.1% formic acid (FA). Elution was carried out with a linear gradient of 5-35% buffer B in buffer A for 30 min (buffer A: 0.1% FA; buffer B: acetonitrile, 0.1% FA) at a flow rate of 300 nl/min. Peptides were analysed using a LTQ Orbitrap Velos Pro (Thermo Scientific) operating in information-dependent acquisition mode using a top 15 method. MS spectra were acquired in the Orbitrap analyser with a mass range of 335-1800  $m/z$ , with a resolution of 60,000 in the Orbitrap. Collision-induced dissociation (CID) peptide fragments were acquired in the ion trap with a collision energy of 35, activation energy of 0.25 and 10 ms activation time, with a default charge state of 2 for fragment ions. Orbitrap Velos RAW data files were extracted and converted to Mascot generic files (.mgf) for database searching using Mascot v2.4.1 (Matrix Science, London, UK).

#### *Database searching*

All MS/MS samples were analysed using Mascot (Matrix Science, London, UK; version 2.4.1). Mascot was set up to search the LIMS12524\_20171212 *C. daubneyi* transcriptome database (version 1.0, 48899 entries) assuming the digestion enzyme strict trypsin with 1

missed cleavage allowed. The transcript data can be accessed via the European Nucleotide Archive ([www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena)) under accession number PRJEB28150. Mascot was searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 10.0 parts per million. Carbamidomethylation of cysteine was specified in Mascot as a fixed modification. Gln->pyro-Glu of the N-terminus, deamidation of asparagine and glutamine, oxidation of methionine, dioxidation of methionine and acetylation of the N-terminus were specified in Mascot as variable modifications.

#### *Criteria for protein identification and quantitation*

Scaffold (version Scaffold\_4.4.5, Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (35). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Additionally, a label-free quantitative analysis was performed in Scaffold for those proteins (NEJ vs adult), with at least two unique peptides, that were present in all three biological replicates. The exponentially modified protein abundance index (emPAI) was used as quantitative method with a t-test (Benjamini-Hochberg FDR correction; significance level,  $p < 0.05$ ) as statistical method. For quantitation, missing values were replaced with a minimum value (default of zero) and normalization was performed with zero as minimum value.

#### *Coproantigen ELISA*

The coproantigen ELISA was tested against faecal samples taken from cattle naturally-infected with *C. daubneyi* and with faecal egg counts ranging from 12 - 299 eggs per gram. Adult E/S products were used as a positive control. Faecal supernatants were prepared using a method adapted from Teimoori *et al* (36). Briefly, 1 g faeces were homogenised in 3 mL lysis buffer (20 mM Tris-HCl, 0.5% SDS, 8M urea) and rotated overnight at room temperature (18-21°C). The extracts were centrifuged at 8000 x g for 10 min, with the supernatant further centrifuged at 10,000 x g for 5 min and the final supernatant stored at -20°C until use. Affinity-purified antibodies ( $\alpha$ -CdE/S) were raised in rabbit against whole E/S proteins collected from adult *C. daubneyi* maintained *in vitro* (EuroGenTec). A proportion of these antibodies were biotinylated ( $\alpha$ -CdE/S-B). 96 well flat bottom microtiter plates (Immulon 2HB, Thermo Scientific) were incubated with 50  $\mu$ l (2  $\mu$ g/ml)  $\alpha$ -CdE/S Ig diluted in PBS per well and incubated overnight at 4°C. Plates were washed six times in PBS containing 0.5 % Tween 20, then blocked for 1 h at room temperature in 1% skimmed milk/PBS-Tween. Fifty  $\mu$ l of the faecal supernatants were then added, at a dilution of 1:27 in PBS, and incubated for 2 h at room temperature. Plates were washed as before and 50  $\mu$ l (2  $\mu$ g/ml)  $\alpha$ -CdE/S-B added per well and incubated for 1 h at room temperature. Plates were washed as before and 50  $\mu$ l of ExtrAvidin-alkaline phosphatase conjugate (Sigma) (1:1500 dilution) was added per well. Plates were washed as before prior to the addition of 100  $\mu$ l per well of SigmaFast p-nitrophenyl phosphate substrate (Sigma-Aldrich). After incubation at room temperature for 30 min plates were read at 405 nanometers (nm) using a Polarstar Omega micro-plate reader (BMG Labtech).

For assessment of specificity, 96-well plates were coated with 2  $\mu$ g/ml of whole worm extracts (WWE) or E/S products (50  $\mu$ l/well) from *F. hepatica* and a range of gastrointestinal nematode parasites that are commonly found in ruminants in the UK/Ireland (*Teladorsagia circumcincta*, *Ostertagia ostertagi*, *Haemonchus contortus*, *Trichostrongylus axei* and

*Trichostrongylus vitrinus*) and incubated overnight at room temperature. Plates were washed and blocked as described above and then 2 µg/ml of α-CdE/S (50 µl/well) was applied for 2 h at room temperature. Plates were washed as before and 50 µl per well of anti-rabbit IgG/alkaline phosphatase conjugate (Sigma) applied at a 1:1500 dilution. Plates were washed, developed and read as described above. For all ELISAs, samples were analysed in triplicate and results are the representative of at least three independent experiments. ELISA results were considered positive for rumen fluke infection if returning an optical density greater than the mean plus two standard deviations of uninfected control samples. This cut-off was also used to define no cross-reactivity with the other species tested.

#### *Immunofluorescence microscopy and whole mount preparations*

Adult *C. daubneyi* and *F. hepatica* were fixed with 4% paraformaldehyde in PBS (Sigma-Aldrich) overnight at 4 °C and subsequently embedded in JB-4 resin (Sigma-Aldrich). Semi-thin sections, 2 µm thick, were cut on a pyramitome and mounted on clean glass slides. For immunofluorescence, JB-4 sections were washed with PBS and then incubated in 10 µg/ml of affinity-purified polyclonal α-CdES in antibody diluent (AbD: PBS containing 0.2% (v/v) Triton X-100) overnight at 4 °C. As a negative control, comparable sections were incubated in rabbit pre-immune serum. The sections were then washed three times in AbD before incubation in a 1:100 dilution of the secondary antibody, fluorescein isothiocyanate (FITC)-conjugated goat anti-rabbit IgG (Sigma-Aldrich), in AbD for 1 h at room temperature. Following three washes in PBS, the sections were mounted in glycerol containing 10% (v/v) PBS and 0.1 M propyl gallate (Sigma-Aldrich) then viewed under a Leica DM2500 fluorescent microscope.

For whole mount preparations, flukes were flat-fixed in formol acetic alcohol solution for 3-5 days at room temperature then incubated in 1% (w/v) carmine (Sigma) for 1.5 hours

at room temperature. Flukes were then de-stained in 70% acidified ethanol overnight and dehydrated in a series of alcohols. The samples were then incubated overnight with HistoChoice clearing agent (Sigma) and mounted on glass slides using DPX mounting media (Sigma). Flukes were viewed using a Leica DM2500 microscope.

## Results

### *The intra-mammalian life-cycle stages of C. daubneyi*

The intra-mammalian life-cycle of *C. daubneyi* involves four major developmental stages: 1) the NEJ flukes that emerge from ingested metacercariae and migrate into the duodenal submucosa; 2) the immature worms that actively feed on host tissue in the intestine; 3) small newly-migrated flukes that have recently completed their migration to the rumen from the duodenum and 4) mature adult flukes that are well-established in the rumen. Macroscopic/microscopic examination of these stages shows their developmental progression (Figure 1). The NEJs are typically 150 µm in length and already bear the oral and posterior suckers characteristic of paramphistomes. Histological examination revealed that whilst no reproductive structures were present, sub-tegumental muscle bundles and a developing bifurcated gut were observed (data not shown). The immature flukes (2-3 mm in length) are generally found in association with the duodenal mucosa though some remain within the intestinal lumen in very heavy infections. A prominent bifurcated gut was clearly visible in carmine-stained whole mount worms as were genital anlagen corresponding to rudimentary testes (anterior and posterior) and ovary. No vitellaria was observed at this stage (Figure 1).

The newly-migrated fluke (typically 5-7 mm in length) are found in the rumen attached to the rumen papillae, wall or contents. Live worms are bright red in appearance and share similar morphology to adult fluke but are much smaller in size. In whole mount preparations the testes and ovary appear larger and more developed, and vitelline follicles

have now appeared around the periphery of the fluke body. Despite their small size, eggs were often observed in the uterus of newly-migrated flukes (Figure 1). The eggs produced by newly-migrated *C. daubneyi* were viable and released highly motile miracidia during *in vitro* egg hatch assays performed as previously described (37). Adult *C. daubneyi* (typically 1.5-2.0 cm in length) appear mid to pale pink with redder colouration around both the anterior and posterior suckers. They are found attached to the rumen wall or papillae via their posterior sucker. Whole mount adult flukes showed fully developed reproductive structures with the posterior testis and egg-filled uterus being the most prominent features (Figure 1).

#### *The C. daubneyi transcriptome*

The PacBio Iso-Seq approach generated a final dataset of 27,524 high quality transcript sequences. Summary statistics for the Iso-Seq transcriptome sequencing and functional annotation are shown in Table 1. Functional annotation of the final 27,524 transcripts was performed using the Trinotate pipeline (20). A total of 17,793 transcripts (68.55%) with a topstrand protein prediction returned an assigned annotation using Blast X to interrogate the TrEMBL database. The majority of annotations returned were from related trematodes, with some hits from cestode and nematode species also. Only 4.1% of sequences returned an annotation to organisms outside of these taxonomic groups, whilst 7,168 sequences could not be annotated. Annotated predicted topstrand protein sequences were grouped according to their eggNOG functional class, with intracellular trafficking/secretion/vesicle transport and post-translational modifications/protein turnover/chaperones classifications being particularly abundant (Figure 2).

#### *Expression patterns of transcript functional groups are associated with parasite development*



Mapping of the RNAseq Illumina reads to the full length (Iso-seq) transcriptome facilitated differential expression (DE) analysis across the four intra-mammalian lifecycle stages. Principle component analysis showed that there was clustering of the biological replicates but clear separation between each developmental stage (Supplementary Figure S1). An overview of the number of differentially expressed transcripts identified between each lifecycle stage is shown in Supplementary Table S2. To identify transcripts with the greatest levels of differential expression across the four life cycle stages, those with a logFC value of at least  $\pm 9$  in at least one pair-wise comparison across the whole dataset were selected. These sequences could be assigned to 10 distinct clusters based on their expression patterns (Figure 3). The annotation for the top 10% DE transcript sequences is shown in Supplementary Table S1. Notably, cysteine peptidases appeared in 4 of the 10 clusters, with each cluster having a different expression pattern suggesting developmental regulation of individual enzymes (see below). Aspartic and cysteine peptidases as well as vitelline proteins were present in cluster 1, which is comprised of transcripts that are up-regulated in the egg-producing newly-migrated and adult stages. Transcripts encoding structural/cytoskeletal proteins including annexin, dynein and tubulin were present in cluster 2, up-regulated in the adult flukes. Transcripts encoding fatty acid-binding proteins were assigned to clusters 2, 4 and 5 and showed up-regulation in the newly migrated and adult stages. Oxygen-scavenging myoglobins were represented in clusters 3, 7 and 10. In the latter group, transcripts encoding serine proteases and serine protease inhibitors were prevalent, indicating their up-regulation in the NEJ and immature stages of *C. daubneyi*. Cysteine-rich secretory proteins (CRSPs) were also prominent in clusters 2 and 3 with up-regulation in adult and newly-migrated stages respectively.

Major metabolic pathways used by parasitic helminths also showed considerable changes in expression across the intra-mammalian stages of *C. daubneyi*. Initial BLAST

analysis revealed that transcripts encoding key enzymes of the glycolysis/gluconeogenesis, tricarboxylic acid cycle (TCA) and malate dismutation pathways were conserved in *C. daubneyi*. We found that most members of the glycolysis/gluconeogenesis pathway showed highest expression in the adult flukes. In contrast, the majority of TCA cycle members were expressed more highly by NEJ/early stages. Finally, expression of transcripts encoding members of the malate dismutation pathway gradually increased with fluke development and showed significant up-regulation within the adult stage (Supplementary Figure S2).

### *C. daubneyi* expresses orthologues of helminth immunomodulators during acute and chronic infection

A defining feature of helminths is their ability to modulate the host immune response to promote their longevity and reproduction. Whilst many individual molecules with reported immunomodulatory activity have been identified from a range of trematode species, we know very little about the secretion of such molecules by paramphistomes. To begin to address this, we used the sequences of known trematode immunomodulators as queries for BLASTP analysis to determine whether orthologues were expressed by *C. daubneyi*. BLAST searches using the trematode-specific helminth defence molecules (HDMs; 38) as queries identified 13 transcripts. Phylogenetic analysis showed that these sequences separated into 3 distinct clades, each containing/encoding one dominant isoform (based on TPM values). Members of these clades were termed CdHDM-1,-2 and -3 respectively (Supplementary Figure S3). To determine the expression pattern for the HDM orthologues across the *C. daubneyi* life stages, the sum of the TPM values for members of each clade (1, 2 and 3) were calculated. This revealed that HDM transcripts were more highly expressed in the NEJ and immature stages

than in the rumen-dwelling newly-migrated and adult stages with the exception of clade 2 CdHDMs which were also upregulated in the newly-migrated stage.

Proteins encoded by members of the CdHDM transcript clades were also identified in the secretomes of NEJs and adult flukes by mass spectrometry. Although CdHDM-2 and -3 showed lowest transcription in the adult flukes compared to the other lifecycle stages, a CdHDM-3 clade member was found to be the most abundant protein detected in adult *C. daubneyi* E/S when analysed by LC-MS/MS. Indeed, both CdHDM-2 and -3 proteins were detected in adult and NEJ E/S preparations. Helical wheel analysis indicated that only CdHDM-2 and -3 contained the conserved C-terminal amphipathic helix, which is a key functional feature of this group of immunomodulators (38 – 40) (Supplementary Figure S3).

Orthologues of other trematode immunomodulators, including kunitz-type inhibitor, peroxiredoxin, fatty acid-binding proteins (FaBP), venom allergen-like (VAL) 9/13 and glutathione *S*-transferase (GST) sigma were also identified following BLAST analysis. At the transcript level, most of these molecules showed highest expression in *C. daubneyi* NEJs whilst some (e.g. GST sigma and FABP 2/15) showed highest expression in the immature and/or newly-migrated stage (Figure 3).

#### *Qualitative and quantitative analysis of the secretome of C. daubneyi NEJs and adult flukes*

Excretory-secretory (E/S) proteins released by *C. daubneyi* NEJs and adult flukes during *in vitro* cultivation were concentrated from culture supernatants and analysed by LC-MS/MS against our *C. daubneyi* Iso-Seq transcriptome. For this analysis, a protein match was only accepted if all 3 biological replicates contained 2 or more unique matched peptides. A total of 506 proteins were identified in the NEJ E/S and 540 proteins in the adult E/S after applying these stringent criteria (Supplementary Table S3). Of these, 272 proteins were unique to the

NEJ secretions, 306 unique to the adult secretions and 234 proteins identified in both samples.

The top 50 most abundant proteins (based on emPAI scores) in the NEJ and adult E/S samples were grouped into functional categories (Figure 4). In the NEJ secretome, proteins with binding functions were particularly abundant, representing nearly 40% of the identified proteins. This group included several myoglobins, which act as oxygen scavengers, and a number of calcium-binding proteins. Approximately 5% have roles in defence, such as the GSTs and CdHDM-2. Metabolic/other enzymes and structural (cytoskeletal and membrane) proteins comprised a large proportion of the top 50 most abundant proteins, as did a number of uncharacterised proteins.

In the adult secretome, defence-associated proteins were especially abundant, representing over 35% of the total proteins in this group. As stated above, a CdHDM-3 clade member was the single most abundant protein based on emPAI value. Other molecules categorised as 'defence' belonged to the GST family. A smaller proportion of the top 50 most abundant adult E/S proteins have roles in binding, and other functional groups are of similar proportions to those seen in the NEJ secretions.

A label-free quantitative approach (based on emPAI values) was used to determine the respective levels of the proteins secreted by *C. daubneyi* NEJs and adult flukes. Volcano plot representation of the data identified several proteins such as serpins, myoglobins and legumain that were significantly ( $p < 0.05$ ) enriched in the NEJ secretome (Figure 5). These molecules were also members of the NEJ-associated transcript group 10 (Figure 3) confirming their expression/secretion by NEJs at the protein level. In contrast, aspartic proteases, cathepsin B peptidases and myoglobin (transcript group 1; Figure 3), annexins and fatty-acid binding proteins (transcript group 2; Figure 3) were significantly enriched ( $p < 0.05$ ) in the adult secretome (Figure 5). Other proteins enriched in the adult fluke secretome,

with potential roles in defence, included GST, thioredoxin and peptidoglycan-recognition protein. These results show that *C. daubneyi* secretory proteins are developmentally regulated and correlate with the migration of the fluke within the gastrointestinal tract of the ruminant host.

The top 50 most abundant proteins in the NEJ and adult *C. daubneyi* secretomes (as described above) were compared with the top 50 proteins secreted by *F. hepatica* NEJs and adult flukes. *F. hepatica* was chosen for comparative analysis as it is frequently found co-infecting the same ruminant host as *C. daubneyi* and because the secretions from both species were collected at the same time-points following *in vitro* culture; 24 h for NEJs (41) and 5 h for adult flukes (42). At the NEJ stage there was a 24% (12/50) overlap with actin, cathepsin B, enolase, fasciclin-1, FaBP, fructose-bisphosphate aldolase, GST, myoglobin, nucleoside diphosphate kinase, protein disulphide isomerase, serpin and superoxide dismutase present in both species. For the adult stages, there was a 20% (10/50) overlap with annexin, cathepsin B, enolase, FaBP, fructose-bisphosphate aldolase, GST, HDM, legumain, myoglobin and triosephosphate isomerase detected in the secretome of both species.

#### *Cysteine peptidase families have undergone selective expansion in C. daubneyi*

Cathepsin-like cysteine peptidases dominate the transcriptome and secretome of many trematode species (43, 44). Accordingly, we used the primary amino acid sequences of cathepsins B, D, F and L from *Schistosoma mansoni*, *Opisthorchis viverrini*, *Fasciola hepatica* and *Clonorchis sinensis* as queries for BLAST analysis of our *C. daubneyi* transcriptome. Using this approach, a single transcript (Cdaub\_01253) encoding a 53.0 kDa cathepsin F-like cysteine peptidase was identified. InterPro analysis showed that the molecule possessed the N-terminal proteinase inhibitor (prosegment) domain (IPR013201)

and the C-terminal C1A peptidase domain (IPR000668) which are typical of cysteine peptidases. An additional cystatin domain (SSF5440) was present N-terminal to the prosegment and contained a partial QXVXG motif (QAPGG) required for inhibition of cysteine peptidases (45). In the prosegment region, the ERFNAQ-like and E/DXGTA motifs characteristic of cathepsin Fs (46, 47) were present whilst the GXNFXFD-like motif required for pH-dependent auto-processing of cysteine peptidases (48) was partially conserved (GITPFSD).

A further two transcripts (Cdaub\_01489 and Cdaub\_08378) encoding cathepsin L-like cysteine peptidases were also identified although the latter was truncated at the 5' end. For the full length sequence, InterPro analysis confirmed the presence of the C-terminal peptidase domain (IPR000668) and the N-terminal prosegment (IPR013201). Although the GXNFXFD-like motif required for pH-dependent processing was fully conserved (GVNEFSD), conserved asparagine residues found at the boundary of the prosegment/mature peptidase domain (allowing initial *trans*-activation by asparaginyl endopeptidase/legumain) in other trematode cathepsin Ls (49) were absent. Furthermore, a specific LSH cleavage site required (in *F. hepatica*) for subsequent removal of the prosegment of cathepsin L zymogens by *trans*-activated cathepsin L molecules (50) was not conserved in the *C. daubneyi* cathepsin L sequence. The S2 sub-site, a binding pocket that governs substrate specificity of cysteine peptidases, is composed of residues occupying positions 67, 68, 133, 157, 160 and 205 with positions 67, 157 and 205 being particularly important (51). The key S2 sub-site residues were found to differ between the two *C. daubneyi* cathepsin L-like peptidases: Cdaub\_01489 (S2 sub-site: Leu<sup>67</sup>, Leu<sup>157</sup> and Leu<sup>205</sup>) and Cdaub\_08378 (S2 sub-site: Tyr<sup>67</sup>, Met<sup>157</sup> and Ala<sup>205</sup>) indicating that they will have distinct substrate preference.

In contrast to the cathepsins F and L, BLAST analysis found significant expansion of *C. daubneyi* transcripts encoding cathepsin B and aspartic proteases (cathepsin D). A total of

27 cathepsin B transcripts were identified. A phylogenetic analysis of the *C. daubneyi* cathepsin B protein sequences revealed that they separated into six well-supported clades (Figure 6A). Mapping of TPM values against the transcripts showed that the clade members are developmentally expressed with clades 1, 3A and 5 showing highest expression in the intestinal NEJ and immature stages whilst members of clades 2, 3B, 4, 6A and 6B are expressed predominantly by rumen-dwelling newly-migrated and adult flukes (Figure 6B). As for cathepsin L cysteine peptidases, the substrate specificity of cathepsin Bs is also determined by the composition and arrangement of amino acids that create the S2 active site sub-site. Using primary sequence alignments and analysis of the atomic structure of *S. mansoni* cathepsin B, the key residues in this pocket that interact with the P2 amino acid of the substrate have been identified as those situated at positions 100, 146, 244, 245, 269 and 316 (52). A comparison of the amino acids that occupy these positions in the various phylogenetic clades of the *C. daubneyi* cathepsin B family is shown in Table 2 and reveals a number of substitutions that could have a critical impact on their substrate preferences.

Similarly, a family of 28 aspartic proteases were identified following BLAST analysis of the *C. daubneyi* transcriptome. Phylogenetic analysis segregated the sequences into 7 clades (Supplementary Figure 4A) which again displayed developmental regulation: whilst expression was low for all aspartic proteases in the NEJ stage, clade 3 enzymes and two members of clade 7 showed highest expression in the immature flukes. The remaining clade members showed clear up-regulation in the newly-migrated and adult worms (Supplementary Figure 4B).

At the protein level, cysteine peptidases were identified in the secretions of *C. daubneyi* NEJs and adult flukes although, based on emPAI values, they did not appear to be the dominant molecules: only 2 out of the 6 cathepsin Bs identified in the NEJ secretions were in the top 50 most abundant proteins. For the adult secretome, it was 5 out of 15.

Similarly, none of the 2 aspartic proteases identified in the NEJ secretions were in the top 50 proteins and for the adult secretome it was 1 of 11. The cathepsin F, and both cathepsin L peptidases, were identified in the secretome of *C. daubneyi* NEJs whilst a single cathepsin L (Cdaub\_01489) was also present in adult fluke secretions (none were present in the list of top 50 proteins).

#### *Development of a coproantigen-based ELISA for detection and diagnosis of C. daubneyi*

A coproantigen-based sandwich ELISA was developed to assess the potential of E/S for diagnosis of *C. daubneyi* infection in ruminant livestock. Anti-Cd-E/S antibodies raised in rabbits were coated on ELISA plates and used to capture *C. daubneyi* antigens present in faecal supernatants from naturally infected cattle. This approach gave high sensitivity of diagnosis (93%), with 28/30 samples from infected animals showing reactivity greater than 2 standard deviations of the mean uninfected absorbance (Figure 7A). To determine the limit of detection of the assay, adult *C. daubneyi* E/S products were serially diluted (from 3 mg/ml) and tested in the antigen-capture ELISA. Taking the limit of detection to be two standard deviations above the mean background absorbance, regression against the linear portion of the dilution series yielded a limit of detection of  $7.0 \pm 0.17$  ng/ml. The specificity of the assay was assessed by ELISA by incubating the  $\alpha$ -Cd-E/S antibodies with E/S or WWE from the liver fluke, *F. hepatica*, and a number of gastrointestinal nematode parasites commonly found in ruminants in the UK/Ireland (*T. circumcincta*, *O. ostertagi*, *H. contortus*, *T. axei* and *T. vitrines*). No cross-reactivity was observed (Figure 7B).

Immunofluorescence was used to determine the origin of the secreted antigens recognised by the  $\alpha$ -Cd-E/S antibodies (Figure. 8). Sections of adult *C. daubneyi*, probed with the  $\alpha$ -Cd-E/S antiserum, showed strong immunolabelling of the tegumental syncytium and



underlying tegumental cell bodies. Intense immunoreactivity was also observed in the lamellae that project from the gastrodermal cells that line the fluke gut. In contrast, the  $\alpha$ -Cd-E/S antiserum did not recognise equivalent tissues from *F. hepatica*, corroborating the ELISA results. No immunolabelling was observed when sections were probed with pre-immune sera.

## Discussion

Despite the rise in prevalence of *C. daubneyi* across much of Western Europe (1) and the sporadic outbreaks of acute, fatal, paramphistomosis (e.g. 4) very little is known about the basic biology and pathogenicity of this species. To begin to address this, we have produced the first transcriptome datasets for four intra-mammalian life-cycle stages of *C. daubneyi* and characterised the secretome of the NEJ and adult flukes that are responsible for acute and chronic disease respectively. In doing so we have provided the first description of how various metabolic, virulence and invasive factors are regulated with parasite development and have used these molecules to develop the first ELISA-based coproantigen test for diagnosis of *C. daubneyi* infection of livestock.

Transcriptome data is currently available for two paramphistome species, *Paramphistomum cervi* and *C. daubneyi* (53, 54). However, these studies have been limited to the adult life-cycle stage which are often described as “well-tolerated” by their ruminant hosts since they are responsible for chronic paramphistomosis which is largely subclinical (13). Here we present a comparative transcriptome analysis of the adult stage with the previously uncharacterised NEJs (infective stage), immature larvae and newly-migrated flukes (migratory stages). We initially used the PacBio platform to generate full length transcripts for all four life-cycle stages without the need for troublesome assembly of short reads. PacBio sequencing identified a total of 24,956 full length transcripts with a top-strand protein coding region. This method almost certainly avoids transcript overestimation, with

the number of transcripts described here for *C. daubneyi* approximately half of those reported for other trematode parasites using *de novo* assembly (e.g. 53 – 5) but similar to those which have been produced using genome-guided approaches (57 – 59).

Changes in gene transcription were followed across the four intra-mammalian developmental stages of *C. daubneyi* using TPM values derived from RNA-seq libraries (in biological triplicate for each stage). Principal component analysis revealed clustering of the replicates but clear separation between life-cycle stages. This demonstrates that each life-cycle stage display distinct patterns of temporal gene expression which correlate with both their development/maturity and niche within the mammalian host. Differential expression analysis revealed that the most highly regulated transcripts (i.e. those showing at least a 9-log fold change in at least one comparison between stages) were dominated by a few functional groups. Some of the most striking transcriptional changes involved up-regulation in both the newly-migrated and adult stages. These included tubulins and dynein, both key structural components of the axoneme of trematode spermatozoa (60,61) and the vitelline proteins which comprise trematode eggshells (62). This temporal expression pattern of reproduction-associated genes follows the appearance of the reproductive organs during *C. daubneyi* development with the anterior and posterior testes, ovary and vitellaria first seen (all together) by the newly-migrated stage. Despite the clear size difference between the newly-migrated and adult rumen fluke, both stages produced eggs. Thus, it is likely that egg production begins rapidly upon the arrival of the migrating flukes into the rumen, similar to the almost immediate onset of egg production observed when juvenile *F. hepatica* enter the bile duct from the liver parenchyma (62).

A total of 7.5% of our predicted protein sequences possessed an N-terminal signal peptide for secretion via the endoplasmic reticulum/Golgi pathway which is comparable to other trematodes such as *C. sinensis* (6.5%) and *O. viverrini* (6.9%) (56). Indeed, the majority

of the most highly regulated transcripts encoded proteins found in *C. daubneyi* secretions following LC-MS/MS analysis. Moreover, our quantitative proteomics analysis showed that many proteins including myoglobins, serpins, cysteine peptidases, aspartic peptidases, GSTs and FaBPs were developmentally regulated which corroborates the transcriptome analysis. The secretomes of two key stages in the *C. daubneyi* intra-mammalian life-cycle were analysed: the infectious NEJs which are responsible for acute disease in the duodenum and the mature adult worms found in the rumen during chronic infection. Proteins with binding function dominated the secretions of both stages, notably 14-3-3 proteins, calcium-binding proteins, FaBPs, ferritin and myoglobin isoforms. Although some of these molecules have reported immunomodulatory functions in other trematodes (63 – 65), their capacity for binding would also allow them to function as carriers to facilitate uptake of various molecules from the host microenvironment *in vivo*. Interestingly, ferritin was enriched in adult fluke secretions compared to those of the NEJ. Ferritin is believed to act as an iron carrier following extracellular haemoglobin digestion in the schistosome gut (66). Since adult *C. daubneyi* are not thought to feed directly on host blood (due to the absence of supporting pathology) (1) secreted ferritin may sequester iron from rumen fluid or have alternative immunomodulatory roles as described for *Clonorchis sinensis* (67). Similarly, myoglobin was highly expressed by *C. daubneyi* at the transcript level and different isoforms were found to be enriched in NEJ and adult fluke secretions. Myoglobins act as oxygen scavengers and are thought to be important for parasite metabolism through oxygen transport/storage (68, 69). Trematode myoglobins have some of the highest recorded affinities for oxygen (70) which would be of particular value in the largely anaerobic rumen (71).

Metacercariae are non-feeding stages and must rely upon endogenous glycogen stores to support them during excystment and initial establishment of infection. In *F. hepatica* the relative expression levels of key enzymes that regulate glycogen synthesis and degradation

indicate that metacercarial glycogen stores are rapidly used up following excystment (41). Similarly, fructose-1,6-bisphosphatase (involved in glycogen synthesis) and phosphofructokinase (involved in glycogen degradation) were both up-regulated in *C. daubneyi* NEJs indicating that the fluke is expending pre-existing glycogen stores whilst simultaneously replenishing them during early infection. Conceivably, this provides the NEJ with sufficient energy to establish infection (e.g. to support active secretory processes and locomotion required during penetration into the duodenal mucosa) whilst creating new reserves to metabolise during the long migration towards the rumen.

Trematodes undergo drastic changes in oxygen availability as they progress through their life-cycle. Metacercariae, and other free-living stages, readily obtain oxygen from the environment. However, once infections become established and flukes migrate within host tissues they begin to encounter environmental hypoxia (72, 73) which coincides with a switch from aerobic energy metabolism (TCA cycle) to anaerobic malate dismutation (74). Our transcriptome analysis indicates that *C. daubneyi* follows this metabolic strategy with TCA cycle members expressed most highly by the NEJ stage and becoming gradually down-regulated as the parasite matures to the adult stage. Coincident with the steady decline of the TCA pathway is the up-regulation of members of the malate dismutation pathway with parasite development. Such gradual change of metabolic pathways could suggest the existence of multiple populations of mitochondria in *C. daubneyi* each with different respiratory activities/tissue locations as have been described for the lung fluke, *Paragonimus westermani* (75). This would provide the metabolic plasticity required as *C. daubneyi* moves from aerobic conditions within the duodenal submucosa towards an increasingly anaerobic environment (i.e. the rumen) and ensure that efficient ATP production is maintained to support the enormous energy demands of egg production that begins following this migration. Whilst the biochemistry of *C. daubneyi* mitochondria require further investigation,

Huson et al. (54) did report the production of propionate (a major excretory product of malate dismutation) by adult *C. daubneyi* during *in vitro* culture which supports the current transcriptome data.

Following excystment in the duodenum, NEJs must respond rapidly to adapt to their new environment and to counter inevitable attack by the host immune system (41). At the transcriptional level, *C. daubneyi* orthologues of trematode proteins with known immunomodulatory function including kunitz-type inhibitor, peroxiredoxin, FaBPs, VAL proteins, GSTs were largely up-regulated in the NEJ stage. In *F. hepatica*, and other trematodes, these proteins act, via a variety of mechanisms, to establish a Th2 immune environment which is permissive to parasite survival and reproduction (76 – 81). Whilst the host immune response to *C. daubneyi* infection (and to what extent the parasite can manipulate this) has yet to be characterised in detail, the NEJ stage is clearly equipped with sufficient armoury to avoid the immune response and establish initial infection in the duodenal mucosa. However, field observations, and experimental studies indicate that previous exposure of livestock to rumen fluke provides protection (routinely >99%) against massive infections that would typically cause acute (and often fatal) paramphistomosis (7, 82 – 84). Critically, these studies showed that immunization with adult worms (delivered orally to establish infection directly in the rumen) was not effective and that protective immunity requires exposure of the duodenum to NEJ-derived antigen (7). This is in stark contrast to liver fluke, where there is limited evidence of natural acquired immunity in cattle and less so in sheep (85). As a result, attempts to vaccinate livestock against liver fluke have been largely unsuccessful, despite a considerable research effort (86). Thus, rumen fluke may be more amenable to vaccinal control than liver fluke infection and our proteomics analysis of NEJ secretions has unveiled a range of early-stage molecules which represent potential candidates for vaccine development.

The single most abundant protein found in the adult fluke secretome was CdHDM-3. This represents one of three isoforms of the helminth defence molecules (HDMs), a family of trematode-specific proteins with diverse immunomodulatory functions; *F. hepatica* HDM-1 has been shown to protect mice against lipopolysaccharide (LPS)-induced inflammation by significantly reducing the release of inflammatory cytokines from macrophages (38) and to suppress antigen processing and presentation in macrophages via inhibition of lysosomal vacuolar ATPase (39). This is the first time HDMs have been reported in suborder Pronocephalata and the first time it has been identified in the secretions of *C. daubneyi*, presumably not being resolved on the 2-DE used by Huson et al. (54) due to its size (~ 6 kDa). Whilst the function of HDMs during *C. daubneyi* infection is unknown, two of the isoforms were predicted to possess a C-terminal amphipathic helix which is the key functional determinant of the LPS-binding/immunomodulatory properties of HDMs from other trematodes (38 – 40). Thus, *C. daubneyi* may use HDMs to simultaneously maintain a Th2 immune environment and to protect against pathogen-associated molecular patterns (PAMPs) in the microbiome-rich rumen – both of which would serve to sustain chronic infections which can persist in livestock for many years (7).

In trematodes, cathepsin-like cysteine peptidases play central roles in virulence, infection, tissue migration and modulation of host immune responses (44). Whilst cathepsin B peptidases are expressed by most species, there is clear divergence in the use of other types of cathepsins: *Fasciola* spp. and *Schistosoma* spp. predominantly express cathepsins L, whereas cathepsins F are the major peptidases expressed by *Clonorchis* spp., *Paragonimus* spp. and *Opisthorchis* spp (43). Our transcriptome and phylogenetic analysis showed that the *C. daubneyi* cathepsin B family has undergone considerable expansion and has diverged into six clades with varying S2 subsite residues. Since the composition and arrangement of amino acids that create the S2 subsite within the cathepsin active site determine the specificity of the

enzyme (51, 87) it is likely that the *C. daubneyi* cathepsin B clade members represent a repertoire of enzymes with overlapping/complementary substrate specificities that are capable of degrading a variety of host macromolecules they encounter as they develop and migrate within the host. In support of this the clade members were found, at both the transcript and protein levels, to be developmentally regulated with clades 1, 3A and 5 expressed/secreted predominantly by the NEJs/immature stage and clades 2, 3B, 4 and 6 expressed/secreted by newly-migrated/adult flukes.

*C. daubneyi* was also found to have a large aspartic peptidase family which could be segregated into distinct clades largely expressed by the adult flukes. Despite this, aspartic peptidases were not particularly abundant in the secretome. This mirrors *O. viverrini* which, despite having 27 aspartic peptidase genes, secrete only low levels of the enzyme (59, 88). Although it is not currently possible to predict the active site composition of aspartic peptidases it is likely that, given their expansion and clear developmental regulation, they too will have varying substrate preferences. In blood-feeding helminths, aspartic peptidases usually form part of a multi-enzyme cascade, with endo- and exo-peptidase activity, to degrade host haemoglobin (89). Whilst the source of nutrition for *C. daubneyi* has yet to be confirmed (e.g. host tissue, rumen fluid, rumen microorganisms) such an enzyme cascade is possible since other digestive enzymes including aminopeptidases, carboxypeptidases, chymotrypsin-like peptidases were all found in the secretome. This would be in contrast to *F. hepatica* which relies solely on an expanded repertoire of cathepsin L peptidases, which comprise > 80% of the total soluble protein secreted by adult fluke, to degrade haemoglobin (43). Here, we found only two transcripts encoding cathepsin L (and one encoding cathepsin F) which were secreted by NEJs in low abundance. Whilst these retain key active site residues, indicating that they will form functional enzymes, they lack conserved asparagines the junction of the prosegment and mature domain. This suggests that they are not *trans-*

activated to the mature enzyme by asparaginyl endopeptidase (legumain; although these were also found in *C. daubneyi* secretions) but could be activated by cathepsin B as reported for *O. viverrini* cathepsin F (90).

Due to their abundance in liver fluke secretions, cathepsin L is the target antigen for most coproantigen-based ELISA tests developed for *F. hepatica* infection (91,92). The MM3 monoclonal antibody used for antigen capture in *F. hepatica* coproantigen assays does not cross-react with rumen fluke antigens either on histological sections or by ELISA (93). Our secretome analysis has now confirmed that this is because *C. daubneyi* secrete very little cathepsin L, mostly from the NEJ stage. Thus, while specific tests for fasciolosis are available (which would be of particular value in areas where paramphistomosis also occurs, either singly or as co-infections with *F. hepatica*) there are currently no such tests for rumen fluke infection. As a result of the significant differences in the profile of molecules secreted by both species, we investigated the diagnostic value of *C. daubneyi* E/S antigens in a coproantigen-based ELISA. The results demonstrate that nanogram amounts of *C. daubneyi* E/S antigen could be captured from faecal extracts with sufficient sensitivity to distinguish between naturally-infected and uninfected cattle. Importantly, the anti-CdE/S serum did not cross-react with somatic or secretory antigens from a range of helminth species that commonly share the same ruminant hosts as *C. daubneyi*. Use of this coproantigen ELISA would offer a non-invasive means of diagnosing current rumen fluke infection (antigens only persist for the duration of infection) and could also be used for the detection of drug resistance (94); since current treatment is reliant on a single drug (oxyclozanide) this seems inevitable (1).

In this report we have generated the first transcriptome resources for multiple intra-mammalian life-cycle stages of *C. daubneyi*. By integrating this with secretome analysis of the most clinically-relevant stages, we provide a comprehensive, and dynamic, overview of



initial infection in the duodenum, migration within the gastrointestinal tract and final maturation upon arrival in the rumen. In particular, our analysis of the molecules expressed/secreted by the NEJ stage provides a valuable framework for studies aimed at better understanding infectivity and how drugs or vaccines could be developed to prevent infection and/or limit damaging intestinal pathology associated with acute paramphistomosis. Since pre-patent rumen fluke infections are the primary cause of clinical disease (4) our NEJ secretome data has yielded a range of putative early-stage antigens that could allow timely diagnosis and treatment for this emerging infection.

## **Acknowledgements**

This work was supported by an Industrial Partnership Award (to M.W.R) from the Biotechnology and Biological Sciences Research Council (BB/N017757/1) with additional financial support from Agrisearch and AHDB Beef & Lamb. N.A.M.O. was supported by a postgraduate studentship from the Department for the Economy (DfE) Northern Ireland.

## **Data availability**

The transcriptome data sets supporting the conclusions of this article are available in the European Nucleotide Archive (ENA) repository under accession number PRJEB28150, <http://www.ebi.ac.uk/ena/data/view/PRJEB28150>. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD014550 and 10.6019/PXD014550.

## **References**

1. Huson, K. M., Oliver, N. A. M., Robinson, M. W. Paramphistomosis of Ruminants: An Emerging Parasitic Disease in Europe. (2017). *Trends Parasitol.* **33**, 836–844.
2. Taylor, M. A. Parasite control in sheep: A risky business. (2013). *Small Rumin. Res.* **110**, 88–92.
3. Skuce, P. J., Morgan, E. R., van Dijk, J., Mitchell, M. Animal health aspects of adaptation to climate change: beating the heat and parasites in a warming Europe. (2013). *Animal.* **7**, 333–345.
4. O’Shaughnessy, J., Garcia-Campos, A., McAloon, C. G., Fagan, S., de Waal, T., McElroy, M., Casey, M., Good, B., Mulcahy, G., Fagan, J., Murphy, D., Zintl, A. Epidemiological investigation of a severe rumen fluke outbreak on an Irish dairy farm. (2018). *Parasitology.* **145**, 948–952.
5. Devos, J., Vassiloglou, B., Amenna-Bernard, N., Marcotty, T. Paramphistomosis in sheep; natural infection of lambs by *Calicophoron daubneyi*. (2013). *Rev. Med. Vet. (Toulouse).* **164**, 528–535.
6. Anonymous. Regional veterinary laboratories report. (2016). *Vet. Irel. J.* **7**, 25–30
7. Horak, I. G. Paramphistomiasis of Domestic Ruminants. (1971). *Adv. Parasitol.* **9**, 33–72.
8. Sanabria, R. E. F., and Romero, J. R. Review and update of paramphistomosis. (2008). *Helminthologia.* **45**, 64–68.
9. Fuertes, M., Pérez, V., Benavides, J., González-Lanza, M. C., Mezo, M., González-Warleta, M., Giráldez, F. J., Fernández, M., Manga-González, M.Y., Ferreras, M. C. Pathological changes in cattle naturally infected by *Calicophoron daubneyi* adult flukes. (2015). *Vet. Parasitol.* **209**, 188–196.
10. Mason, C., Stevenson, H., Cox, A., Dick, I. Disease associated with immature

- paramphistome infection in sheep. (2012). *Vet. Rec.* **170**, 343–344.
11. Anonymous. Northern Ireland disease surveillance report, October to December 2016. (2017). *Vet. Rec.* **180**, 112–116.
  12. Foster, A. P., Otter, A., O’Sullivan, T., Cranwell, M. P., Twomey, D. F., Millar, M. F., Taylor, M. A. Rumen fluke (paramphistomosis) in British cattle. (2008). *Vet. Rec.* **162**, 528–528.
  13. Millar, M., Colloff, A., Scholes, S. Disease associated with immature paramphistome infection. (2012). *Vet. Rec.* **171**, 509–510.
  14. Rolfe, P. F., and Boray, J.C. Chemotherapy of paramphistomosis in cattle. (1987). *Aust. Vet. J.*, **64**: 328-332.
  15. Paraud, C., Gaudin, C., Pors, I., Chartier, C. Efficacy of oxclozanide against the rumen fluke *Calicophoron daubneyi* in experimentally infected goats. (2009). *Vet. J.* **180**, 265–267.
  16. Huson, K. M., Wild, C., Fenn, C., Robinson, M. W. Optimized conditions for the in vitro excystment of *Calicophoron daubneyi* metacercariae. (2017). *Parasitology*. **145**, 1015–1019.
  17. Martínez-Ibeas, A. M., González-Warleta, M., Martínez-Valladares, M., Castro-Hermida, J. A., González-Lanza, C., Miñambres, B., Ferreras, C., Mezo, M., Manga-González, M. Y. Development and validation of a mtDNA multiplex PCR for identification and discrimination of *Calicophoron daubneyi* and *Fasciola hepatica* in the *Galba truncatula* snail. (2013). *Vet. Parasitol.* **195**, 57–64.
  18. Li, W., and Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. (2006). *Bioinformatics*. **22**, 1658–1659.

19. Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., Korlach, J. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. (2013). *Nat. Methods*. **10**, 563–569.
20. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N., Regev, A. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. (2013). *Nat. Protoc.* **8**, 1494–1512.
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment search tool. (1990). *J. Mol. Biol.* **215**, 403–410.
22. Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., Bateman, A. The Pfam protein families database: towards a more sustainable future. (2016). *Nucleic Acids Res.* **44**, D279–D285.
23. Eddy, S. R. Accelerated Profile HMM Searches. (2011). *PLOS Comput. Biol.* **7**, e1002195.
24. Petersen, T. N., Brunak, S., von Heijne, G., Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. (2011). *Nat. Methods*. **8**, 785–786.
25. Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Edited by F. Cohen. (2001). *J. Mol. Biol.* **305**, 567–580.

26. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. (2011). *EMBnet.journal*. **17**, 10–12.
27. Langmead, B., and Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. (2012). *Nat. Methods*. **9**, 357-359.
28. Glaus, P., Honkela, A., Rattray, M. Identifying differentially expressed transcripts from RNA-seq data with biological variation. (2012). *Bioinformatics*. **28**, 1721–1728.
29. Love, M. I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. (2014). *Genome Biol*. **15**, 550.
30. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. (1999). *Nucleic Acids Symp. Ser.* **41**, 95-98.
31. Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., Wishart, D. S. Heatmapper: web-enabled heat mapping for all. (2016). *Nucleic Acids Res.* **44**, W147–W153.
32. Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. (2013). *Mol. Biol. Evol.* **30**, 2725–2729.
33. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. (1985). *Evolution*. **39**, 783–791.
34. Zuckerkandl, E., and Pauling, L. Molecules as documents of evolutionary history. (1965). *J. Theor. Biol.* **8**, 357–366.
35. Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. (2003). *Anal. Chem.* **75**, 4646–4658.
36. Teimoori, S., Arimatsu, Y., Laha, T., Kaewkes, S., Sereerak, P., Sripa, M.,

- Tangkawattana, S., Brindley, P. J., Sripa, B. Chicken IgY-based coproantigen capture ELISA for diagnosis of human opisthorchiasis. (2017). *Parasitol. Int.* **66**, 443–447.
37. Fairweather, I., McShane, D. D., Shaw, L., Ellison, S. E., O'Hagan, N. T., York, E. A., Trudgett, A., Brennan, G. P. Development of an egg hatch assay for the diagnosis of triclabendazole resistance in *Fasciola hepatica*: Proof of concept. (2012). *Vet. Parasitol.* **183**, 249–259.
38. Robinson, M. W., Donnelly, S., Hutchinson, A. T., To, J., Taylor, N. L., Norton, R. S., Perugini, M. A., Dalton, J. P. A family of helminth molecules that modulate innate cell responses via molecular mimicry of host antimicrobial peptides. (2011). *PLoS Pathog.* **7**, e1002042.
39. Robinson, M. W., Alvarado, R., To, J., Hutchinson, A. T., Dowdell, S. N., Lund, M., Turnbull, L., Whitchurch, C. B., O'Brien, B. A., Dalton, J. P., Donnelly, S. A helminth cathelicidin-like protein suppresses antigen processing and presentation in macrophages via inhibition of lysosomal vATPase. (2012). *FASEB J.* **26**, 4614–4627.
40. Thivierge, K., Cotton, S., Schaefer, D. A., Riggs, M. W., To, J., Lund, M. E., Robinson, M. W., Dalton, J. P., Donnelly, S. M. Cathelicidin-like helminth defence molecules (HDMs): absence of cytotoxic, anti-microbial and anti-protozoan activities imply a specific adaptation to immune modulation. (2013). *PLoS Negl. Trop. Dis.* **7**, e2307.
41. Cwiklinski, K., Jewhurst, H., McVeigh, P., Barbour, T., Maule, A. G., Tort, J., O'Neill, S. M., Robinson, M. W., Donnelly, S., Dalton, J. P. Infection by the Helminth Parasite *Fasciola hepatica* Requires Rapid Regulation of Metabolic, Virulence, and Invasive Factors to Adjust to Its Mammalian Host. (2018). *Mol. Cell. Proteomics.* **17**, 792–809.

42. Cwiklinski, K., de la Torre-Escudero, E., Trelis, M., Bernal, D., Dufresne, P. J., Brennan, G. P., O'Neill, S., Tort, J., Paterson, S., Marcilla, A., Dalton, J. P., and Robinson, M. W. The extracellular vesicles of the helminth pathogen, *Fasciola hepatica*: Biogenesis pathways and cargo molecules involved in parasite pathogenesis. (2015) *Mol. Cell. Proteomics* **14**, 3258-3273.
43. Robinson, M. W., Tort, J. F., Lowther, J., Donnelly, S. M., Wong, E., Xu, W., Stack, C. M., Padula, M., Herbert, B., Dalton, J. P. Proteomics and Phylogenetic Analysis of the Cathepsin L Protease Family of the Helminth Pathogen *Fasciola hepatica*. (2008). *Mol. Cell. Proteomics* **7**, 1111–1123.
44. Cwiklinski, K., Donnelly, S., Drysdale, O., Jewhurst, H., Smith, D., De Marco Verissimo, C., Pritsch, I. C., O'Neill, S., Dalton, J. P., Robinson, M. W. The cathepsin-like cysteine peptidases of trematodes of the genus *Fasciola*. (2019). *Adv. Parasitol.* **104**, 113–164.
45. Auerswald, E. A., Genenger, G., Assfalg-Machleidt, I., Machleidt, W., Engh, R. A., Fritz, H. Recombinant chicken egg white cystatin variants of the QLVSG region. (1992). *Eur. J. Biochem.* **209**, 837–845.
46. Deussing, J., Tisljar, K., Papazoglou, A., Peters, C. Mouse cathepsin F: cDNA cloning, genomic organization and chromosomal assignment of the gene. (2000). *Gene.* **251**, 165–173.
47. Kang, T. H., Yun, D. H., Lee, E. H., Chung, Y. B., Bae, Y. A., Chung, J. Y., Kang, I., Kim, J., Cho, S. Y., Kong, Y. A cathepsin F of adult *Clonorchis sinensis* and its phylogenetic conservation in trematodes. (2004). *Parasitology*, **128**, 195–207.
48. Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M. C., Storer, A. C., Thomas, D. Y. Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation

- of intramolecular processing. (1995). *J. Biol. Chem.* **270**, 10838–10846.
49. Dalton, J. P., Brindley, P. J., Donnelly, S., Robinson, M. W. The enigmatic asparaginyl endopeptidase of helminth parasites. (2009). *Trends Parasitol.* **25**, 59–61.
50. Stack, C. M., Donnelly, S., Lowther, J., Xu, W., Collins, P. R., Brinen, L. S., Dalton, J. P. The major secreted cathepsin L1 protease of the liver fluke, *Fasciola hepatica*: a Leu-12 to Pro-12 replacement in the nonconserved C-terminal region of the prosegment prevents complete enzyme autoactivation and allows definition of the molecular events in prosegment removal. (2007). *J. Biol. Chem.*, **282**, 16532–16543.
51. Turk, D., Guncar, G., Podobnik, M., Turk, B. Revised definition of substrate binding sites of papain-like cysteine proteases. (1998). *Biol. Chem.* **379**, 137–147.
52. Jílková, A., Řezáčová, P., Lepšík, M., Horn, M., Váchová, J., Fanfrlík, J., Brynda, J., McKerrow, J. H., Caffrey, C. R., Mareš, M. Structural basis for inhibition of cathepsin B drug target from the human blood fluke, *Schistosoma mansoni*. (2011). *J. Biol. Chem.* **286**, 35770–35781.
53. Choudhary, V., Garg, S., Chourasia, R., Hasnani, J. J., Patel, P. V., Shah, T. M., Bhatt, V. D., Mohapatra, A., Blake, D. P., Joshi, C. G. Transcriptome analysis of the adult rumen fluke *Paramphistomum cervi* following next generation sequencing. (2015). *Gene* **570**, 64–70.
54. Huson, K. M., Morphew, R. M., Allen, N. R., Hegarty, M. J., Worgan, H. J., Girdwood, S. E., Jones, E. L., Phillips, H. C., Vickers, M., Swain, M., Smith, D., Kingston-smith, A. H., Brophy, P. M. Polyomic tools for an emerging livestock parasite, the rumen fluke *Calicophoron daubneyi* ; identifying shifts in rumen functionality. (2018). *Parasit. Vectors* **11**, 617.



55. Cantacessi, C., Mulvenna, J., Young, N. D., Kasny, M., Horak, P., Aziz, A., Hofmann, A., Loukas, A., Gasser, R. B. A Deep Exploration of the Transcriptome and “Excretory/Secretory” Proteome of Adult *Fascioloides magna*. (2012). *Mol. Cell. Proteomics* **11**, 1340–1353.
56. Young, N. D., Campbell, B. E., Hall, R. S., Jex, A. R., Cantacessi, C., Laha, T., Sohn, W. M., Sripa, B., Loukas, A., Brindley, P. J., Gasser, R. B. Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*. (2010). *PLoS Negl. Trop. Dis.* **4**, e719.
57. Huang, Y., Chen, W., Wang, X., Liu, H., Chen, Y., Guo, L., Luo, F., Sun, J., Mao, Q., Liang, P., Xie, Z., Zhou, C., Tian, Y., Lv, X., Huang, L., Zhou, J., Hu, Y., Li, R., Zhang, F., Lei, H., Li, W., Hu, X., Liang, C., Xu, J., Li, X., Yu, X. The Carcinogenic Liver Fluke, *Clonorchis sinensis*: New Assembly, Reannotation and Analysis of the Genome and Characterization of Tissue Transcriptomes. (2013). *PLoS One* **8**, e54732 .
58. Protasio, A. V., Tsai, I.J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., de Silva, N., Velarde, G. S., Anderson, T. J. C., Clark, R. C., Davidson, C., Dillon, G. P., Holroyd, N. E., LoVerde, P. T., Lloyd, C., McQuillan, J., Oliveira, G., Otto, T. D., Parker-Manuel, S. J., Quail, M. A., Wilson, R. A., Zerlotini, A., Dunne, D. W., Berriman, M. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. (2012). *PLoS Negl. Trop. Dis.* **6**, e1455.
59. Young, N. D., Nagarajan, N., Lin, S. J., Korhonen, P. K., Jex, A. R., Hall, R. S., Safavi-Hemami, H., Kaewkong, W., Bertrand, D., Gao, S., Seet, Q., Wongkham, S., Teh, B. T., Wongkham, C., Intapan, P. M., Maleewong, W., Yang, X., Hu, M., Wang, Z., Hofmann, A., Sternberg, P. W., Tan, P., Wang, J., Gasser, R. B. The *Opisthorchis viverrini* genome provides insights into life in the bile duct. (2014). *Nat. Commun.* **5**, 4378.

60. Stitt, A. W., and Fairweather, I. Spermatogenesis and the fine structure of the mature spermatozoon of the liver fluke, *Fasciola hepatica* (Trematoda: Digenea). (1990). *Parasitology*. **101**, 395–407.
61. Stitt, A. W., Fairweather, I., Trudgett, A. G., Johnston, C. F., Anderson, S. M. Localisation of actin in the liver fluke, *Fasciola hepatica*. (1992). *Parasitol. Res.* **78**, 96–102.
62. Robinson, M. W., Colhoun, L. M., Fairweather, I., Brennan, G. P., Waite, J. H. Development of the vitellaria of the liver fluke, *Fasciola hepatica* in the rat host. (2001). *Parasitology*, **123**, 509–518.
63. Wang, X., Chen, W., Li, X., Zhou, C., Deng, C., Lv, X., Fan, Y., Men, J., Liang, C., & Yu, X. Identification and molecular characterization of a novel signaling molecule 14-3-3 epsilon in *Clonorchis sinensis* excretory/secretory products. (2012). *Parasitol. Res.* **110**, 1411–1420.
64. Ren, M., He, L., Huang, Y., Mao, Q., Li, S., Qu, H., Bian, M., Liang, P., Chen, X., Ling, J., Chen, T., Liang, C., Wang, X., Li, X., Yu, X. Molecular characterization of *Clonorchis sinensis* secretory myoglobin: Delineating its role in anti-oxidative survival. (2014). *Parasit. Vectors* **7**, 250.
65. Fitzsimmons, C. M., McBeath, R., Joseph, S., Jones, F. M., Walter, K., Hoffmann, K. F., Kariuki, H. C., Mwatha, J. K., Kimani, G., Kabatereine, N. B., Vennervald, B. J., Ouma, J. H., Dunne, D. W. Factors affecting human IgE and IgG responses to allergen-like *Schistosoma mansoni* antigens: Molecular structure and patterns of in vivo exposure. (2007). *Int. Arch. Allergy Immunol.* **142**, 40–50.
66. Hall, S. L., Braschi, S., Truscott, M., Mathieson, W., Cesari, I. M., Wilson, R. A. Insights into blood feeding by schistosomes from a proteomic analysis of worm vomitus. (2011).

*Mol. Biochem. Parasitol.* **179**, 18–29.

67. Mao, Q., Xie, Z., Wang, X., Chen, W., Ren, M., Shang, M., Lei, H., Tian, Y., Li, S., Liang, P., Chen, T., Liang, C., Xu, J., Li, X., Huang, Y., Yu, X. *Clonorchis sinensis* ferritin heavy chain triggers free radicals and mediates inflammation signaling in human hepatic stellate cells. (2015). *Parasitol. Res.* **114**, 659–670.
68. Rashid, A. K., Van Hauwaert, M. L., Haque, M., Siddiqi, A. H., Lasters, I., De Maeyer, M., Griffon, N., Marden, M. C., Dewilde, S., Clauwaert, J., Vinogradov, S. N., Moens, L. Trematode myoglobins, functional molecules with a distal tyrosine. (1997). *J. Biol. Chem.* **272**, 2992–2999.
69. McNulty, S. N., Fischer, P. U., Townsend, R. R., Curtis, K. C., Weil, G. J., Mitreva, M. Systems Biology Studies of Adult Paragonimus Lung Flukes Facilitate the Identification of Immunodominant Parasite Antigens. (2014). *PLoS Negl. Trop. Dis.* **8**, e3242.
70. Kiger, L., Rashid, A. K., Griffon, N., Haque, M., Moens, L., Gibson, Q. H., Poyart, C., Marden, M. C. Trematode hemoglobins show exceptionally high oxygen affinity. (1998). *Biophys. J.* **75**, 990–998.
71. Huang, Y., Marden, J. P., Julien, C., Bayourthe, C. Redox potential: An intrinsic parameter of the rumen environment. (2018). *J. Anim. Physiol. Anim. Nutr. (Berl.)* **102**, 393–402.
72. Müller, M., Mentel, M., van Hellemond, J. J., Henze, K., Woehle, C., Gould, S. B., Yu, R.-Y., van der Giezen, M., Tielens, A. G. M., Martin, W. F. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. (2012). *Microbiol. Mol. Biol. Rev.* **76**, 444–495.
73. McNulty, S. N., Tort, J. F., Rinaldi, G., Fischer, K., Rosa, B. A., Smircich, P., Fontenla,

- S., Choi, Y. J., Tyagi, R., Hallsworth-Pepin, K., Mann, V. H., Kammili, L., Latham, P. S., Dell'Oca, N., Dominguez, F., Carmona, C., Fischer, P. U., Brindley, P. J., Mitreva, M. Genomes of *Fasciola hepatica* from the Americas Reveal Colonization with *Neorickettsia* Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers. (2017). *PLoS Genet.* **13**, e1006537.
74. Tielens, A. G. M. (1999) Metabolism. In: Dalton, J. P., editor. Fasciolosis: Wallingford, CABI Publishing; pp. 277–306.
75. Takamiya, S., Fukuda, K., Nakamura, T., Aoki, T., Sugiyama, H. *Paragonimus westermani* possesses aerobic and anaerobic mitochondria in different tissues, adapting to fluctuating oxygen tension in microaerobic habitats. (2010). *Int. J. Parasitol.* **40**, 1651–1658.
76. Falcón, C. R., Masih, D., Gatti, G., Sanchez, M. C., Motrán, C. C., Cervi, L. *Fasciola hepatica* Kunitz type molecule decreases dendritic cell activation and their ability to induce inflammatory responses. (2014). *PLoS One.* **9**, e114505.
77. Donnelly, S., Stack, C. M., O'Neill, S. M., Sayed, A. A., Williams, D. L., Dalton, J. P. Helminth 2-Cys peroxiredoxin drives Th2 responses through a mechanism involving alternatively activated macrophages. (2008). *FASEB J.* **22**, 4022–4032.
78. Morphew, R. M., Wilkinson, T. J., Mackintosh, N., Jahndel, V., Paterson, S., McVeigh, P., Abbas Abidi, S. M., Saifullah, K., Raman, M., Ravikumar, G., LaCourse, J., Maule, A., Brophy, P. M. Exploring and Expanding the Fatty-Acid-Binding Protein Superfamily in *Fasciola* Species. (2016). *J. Proteome Res.* **15**, 3308–3321.
79. Martin, I., Cabán-Hernández, K., Figueroa-Santiago, O., Espino, A. M. *Fasciola hepatica* fatty acid binding protein inhibits TLR4 activation and suppresses the inflammatory cytokines induced by lipopolysaccharide in vitro and in vivo. (2015). *J. Immunol.* **194**,

3924–3936.

80. Yoshino, T. P., Brown, M., Wu, X. J., Jackson, C. J., Ocadiz-Ruiz, R., Chalmers, I. W., Kolb, M., Hokke, C. H., Hoffmann, K. F. Excreted/secreted *Schistosoma mansoni* venom allergen-like 9 (SmVAL9) modulates host extracellular matrix remodelling gene expression. (2014). *Int. J. Parasitol.* **44**, 551–563.
81. Woo, H. S., Kim, T. Y., Sohn, W. M., Yong, T. S. A tegument-specific venom allergen-like protein of *Clonorchis sinensis*. (2015). *Parasitol. Res.* **114**, 329–333.
82. Boray, J. C. The anthelmintic efficiency of niclosamide and menichlopholan in the treatment of intestinal paramphistomosis in sheep. (1969). *Aust. Vet. J.* **45**, 133–135.
83. Horak, I. G. (1965) Studies on paramphistomiasis v. the immunisation of sheep, goats and cattle preliminary report. *J. S. Afr. Vet. Assoc.* **36**, 361–363.
84. Horak I. G. Host-parasite relationships of *Paramphistomum microbothrium* Fischöder, 1901, in experimentally infested ruminants, with particular reference to sheep. (1967). *Onderstepoort J. Vet. Res.* **34**, 451–540.
85. Spithill, T. W., Smooker, P. M., Sexton, J. L., Bozas, E., Morrison, C. A., Creaney, J., Parsons, J. C. (1999) Development of Vaccines Against *Fasciola hepatica*. In: Dalton J. P., editor. Fasciolosis: Wallingford, CABI Publishing; pp. 377–410.
86. Molina-Hernández, V., Mulcahy, G., Pérez, J., Martínez-Moreno, Á., Donnelly, S., O'Neill, S. M., Dalton, J. P., Cwiklinski, K. *Fasciola hepatica* vaccine: we may not be there yet but we're on the right road. (2015). *Vet. Parasitol.* **208**, 101–111.
87. Lecaille, F., Choe, Y., Brandt, W., Li, Z., Craik, C. S., Brömme, D. Selective inhibition of the collagenolytic activity of human cathepsin K by altering its S2 subsite specificity. (2002). *Biochemistry.* **41**, 8447–8454.

88. Mulvenna, J., Sripa, B., Brindley, P. J., Gorman, J., Jones, M. K., Colgrave, M. L., Jones, A., Nawaratna, S., Laha, T., Suttiprapa, S., Smout, M. J., Loukas, A. The secreted and surface proteomes of the adult stage of the carcinogenic human liver fluke *Opisthorchis viverrini*. (2010). *Proteomics*. **10**, 1063–1078.
89. Williamson, A. L., Lecchi, P., Turk, B. E., Choe, Y., Hotez, P. J., McKerrow, J. H., Cantley, L. C., Sajid, M., Craik, C. S., Loukas, A. A multi-enzyme cascade of hemoglobin proteolysis in the intestine of blood-feeding hookworms. (2004). *J. Biol. Chem.* **279**, 35950–35957.
90. Sripa, J., Laha, T., To, J., Brindley, P. J., Sripa, B., Kaewkes, S., Dalton, J. P., Robinson, M. W. Secreted cysteine proteases of the carcinogenic liver fluke, *Opisthorchis viverrini*: regulation of cathepsin F activation by autocatalysis and trans-processing by cathepsin B. (2010). *Cell Microbiol.* **12**, 781–795.
91. Mezo, M., González-Warleta, M., Carro, C., Ubeira, F.M. An ultrasensitive capture ELISA for detection of *Fasciola hepatica* coproantigens in sheep and cattle using a new monoclonal antibody (MM3). (2004). *J. Parasitol.* **90**, 845–852.
92. Martínez-Sernández, V., Orbegozo-Medina, R. A., González-Warleta, M., Mezo, M., Ubeira, F. M. Rapid Enhanced MM3-COPRO ELISA for Detection of *Fasciola* Coproantigens. (2016). *PLoS Negl. Trop. Dis.* **10**, e0004872.
93. Kajugu, P. E., Hanna, R. E., Edgar, H. W., Forster, F. I., Malone, F. E., Brennan, G. P., Fairweather, I. Specificity of a coproantigen ELISA test for fasciolosis: lack of cross-reactivity with *Paramphistomum cervi* and *Taenia hydatigena*. (2012). *Vet. Rec.*, **171**, 502.
94. Fairweather I., Brennan, G.P., Hanna, R.E.B., Robinson, M.W., Skuce, P.J. Drug resistance in liver flukes. (2020). *Int. J. Parasitol. Drugs Drug Resist.* **12**, 39-59.



## FIGURE LEGENDS

**Figure 1.** The intra-mammalian life-cycle stages of *C. daubneyi*.

(A). Four developmental stages of *C. daubneyi* were subjected to transcriptome analysis. These were: newly excysted juveniles (NEJ) excysted here *in vitro* but typically found in the duodenum *in vivo*, immature flukes found embedded in the duodenal musosa, small newly-migrated flukes flukes just arrived in the rumen and fully developed adult flukes, also found in the rumen. (B). Carmine-stained whole immature, newly-migrated and adult flukes showing the major morphological features and internal organs.

**Figure 2.** Annotation of the *C. daubneyi* transcriptome.

(A) Distribution of transcript annotation hits according to species. Analysis was performed using the top assigned annotation matches from the TREMBL Blast X annotation of all transcript sequences with a topstrand predicted protein sequence (24956). (B) EggNOG functional category annotations of predicted protein sequences returning a NOG/COG identification via the trinotate annotation process (11843/48899 sequences). Categories are broadly grouped into genetic processes (red bars), cell structure and function (blue bars) and metabolism (yellow bars). The largest number of annotations were to the category 'Function unknown' (4765 sequences) which are not shown.

**Figure 3.** Differential expression of *C. daubneyi* transcripts

The top 10% of transcripts showing the most differential expression across our sample set were selected using a cut-off of at least a 9-log fold change in at least one comparison between *C. daubneyi* developmental stages and a minimum average TPM value of 2 in at



least one stage. This returned 2,470 transcripts (Supplementary Table S1), which were grouped into clusters based on their expression pattern. A heatmap of TPM values scaled by transcript values (where red indicates a low value and green a high value) was produced. Individual transcripts were matched to their BlastX annotation from the trinotate output and common proteins/protein family annotations present in each cluster (1-10) are highlighted.

**Figure 4.** Comparative analysis of the secretome of *C. daubneyi* NEJs and adult flukes.

Graphical representation of the composition of the top 50 most abundant secretory proteins (based on emPAI scores) grouped according to function.

**Figure 5.** Qualitative and quantitative proteomics analysis reveals developmental expression of proteins secreted by NEJ and adult *C. daubneyi*.

Proteins identified in the secretome of *C. daubneyi* NEJs and adult flukes (>2 matched peptides in 3 out of 3 biological replicates) were subjected to quantitative analysis shown as a Volcano plot. The x-axis represents  $\log_2$  (fold-change) (adult/NEJ) and the y-axis shows the  $-\log_{10}$  ( $p$  value). The dashed red line indicates the significance threshold ( $p = 0.05$ ). The vertical dashed black line indicates zero fold change. Green boxes represent proteins whose expression is significantly different between the two life stages. The position of selected proteins are shown: CaBP, calcium-binding protein; FaBP, fatty acid-binding protein; GST, glutathione *S*-transferase; LAP, leucine aminopeptidase; PGRP, peptidoglycan-recognition protein; Leg, legumain; Myo, myoferlin; Cat B, cathepsin B; Asp, aspartic protease; Trx, thioredoxin.

**Figure 6.** Phylogenetic relationships of the *C. daubneyi* cathepsin B family and their developmental expression patterns.

(A) Bootstrapped (1000 trials) neighbour-joining phylogenetic tree showing the evolutionary relationship of *C. daubneyi* cathepsin B amino acid sequences. Numbers represent bootstrap values (given as percentages) for a particular node and values greater than 50% are shown. The tree is rooted to *Carica papaya* papain (Uniprot accession P00784). (B) The TPM values of transcripts encoding the cathepsin B clade members were represented as a heatmap to visualise their expression patterns across the intra-mammalian life-cycle stages of *C. daubneyi*. Relative expression is shown by a red to green scale depicting low to high levels of expression, respectively.

**Figure 7.** Confirmation of *C. daubneyi* infection by coproantigen ELISA

A) Faecal samples were prepared from sheep independently determined, by faecal egg count, to be infected or uninfected by *C. daubneyi*. A sandwich ELISA approach was used with native and biotinylated polyclonal antibodies raised against whole adult *C. daubneyi* E/S proteins. Absorbance at 405 nm of infected and uninfected samples. P value from t-test. Hatched line shows absorbance two standard deviations greater than mean uninfected absorbance; red dots indicate samples from infected animals below this limit. B) Specificity of anti-*C. daubneyi* antibodies by ELISA against indicated parasites and life stages.

**Figure 8.** Light micrographs showing  $\alpha$ -CdE/S immunofluorescence in sections of adult *C. daubneyi*.

(A) The tegumental syncytium (T) and underlying tegumental cell bodies (arrows) of *C. daubneyi* show intense and specific labelling with the  $\alpha$ -CdE/S antibody. (B) *C. daubneyi* control section which was incubated in pre-immune rabbit serum. All tissues, including the tegument, are unlabelled. (C) Section showing specific labelling of the lamellae (\*) that arise from the gastrodermal cells (GC) that line the fluke gut. L, gut lumen. (D) *C. daubneyi* control section which was incubated in pre-immune rabbit serum shows no immunofluorescence. (E) and (F) Sections of adult *F. hepatica* show no labelling of the tegument (T) or gastrodermal cells (GC) when probed with the  $\alpha$ -CdE/S antibody. Scale bars represent 25  $\mu$ m.

**Table 1:** Summary statistics for whole transcriptome sequencing and assembly (PacBio IsoSeq) and Trinotate functional annotation of transcript sequences for *C. daubneyi*.

<b>PacBio Sequel Library and Trinotate annotation</b>	
Total transcripts after downstream processing (quality control) and cluster analysis	27,524
Total assembled bases (bp)	26,451,949
Average seq length (bp)	961
Max seq length (bp)	3,452
Min seq length (bp)	114
N50 (bp)	1,128
GC content (%)	45.0
Transcripts with topstrand protein prediction	24,956
Number of annotated topstrand transcripts (Trinotate TREMBL BlastX) (%)	17,793 (68.55)
Proteins predicted (>50 AA)	48,899
Protein seqs with predicted signal peptide (%)	3,696 (7.56)
Protein seqs with predicted transmembrane domain (%)	8,373 (17.12)

**Table 2.** Comparison of the residues from the S2 active site in *C. daubneyi* cathepsin Bs (clades 1-6). Residues were identified using primary sequence alignments and analysis of the atomic structure of *S. mansoni* cathepsin B (52; Protein Data Bank codes 3QSD, 3S3Q, and 3S3R).

	Position					
	100	146	244	245	269	316
<b>Clade</b>						
<b>1</b>	Cys	Pro	Thr	Phe	Ala	Ser
<b>2</b>	Cys	Pro	Asp	Phe	Gly	Asp
<b>3A</b>	Cys	Ser	Ala	Tyr	Gly	Asp
<b>3B</b>	Cys	Pro	Ala	Phe	Gly	Asp
<b>4</b>	Cys	Leu	Gly	Phe	Asn	Phe
<b>5</b>	Cys	Asp	Ser	Leu	Tyr	Arg
<b>6A</b>	Cys	Asp	Ser	Phe	Tyr	Asp
<b>6B</b>	Cys	Asp	Tyr	Phe	Tyr	His

Figure 1.

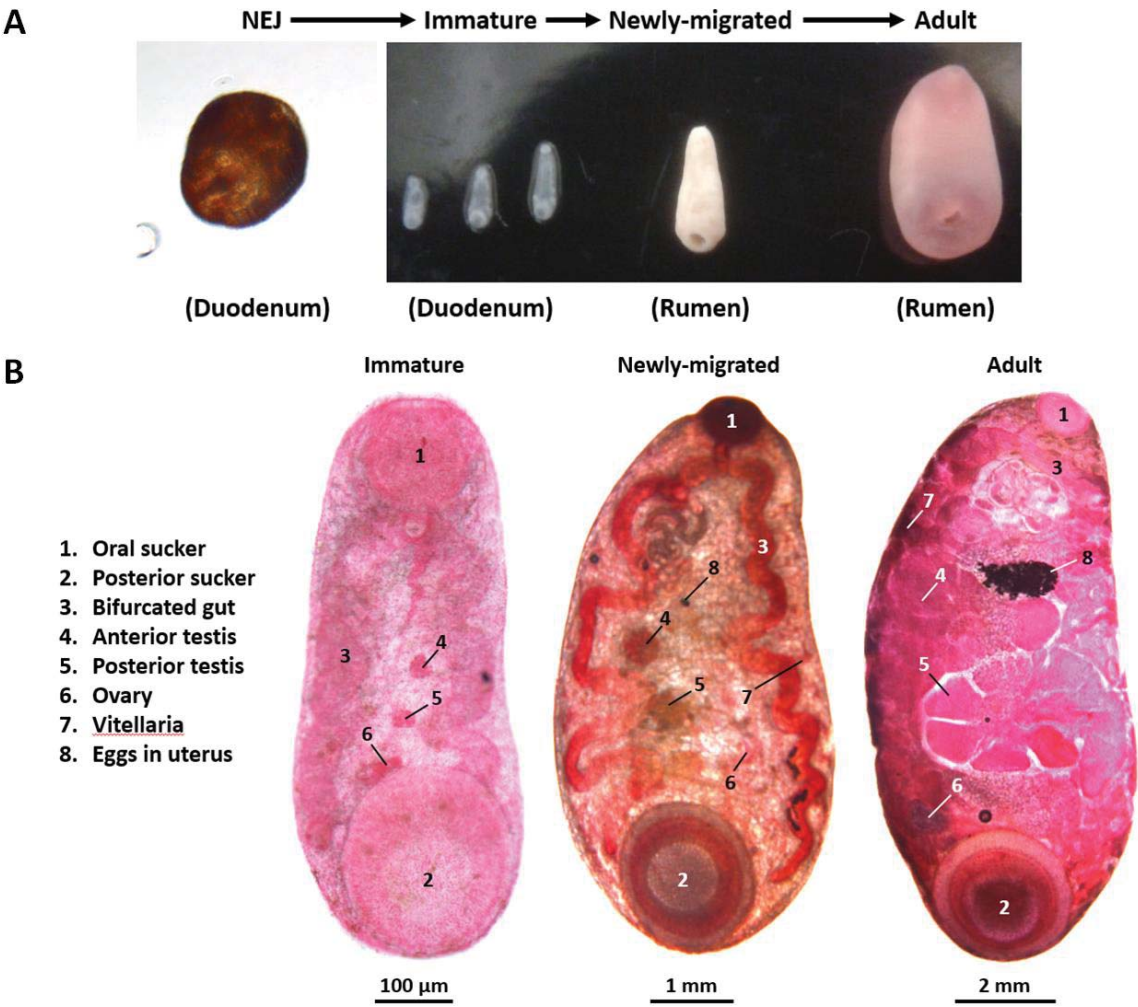


Figure 2.

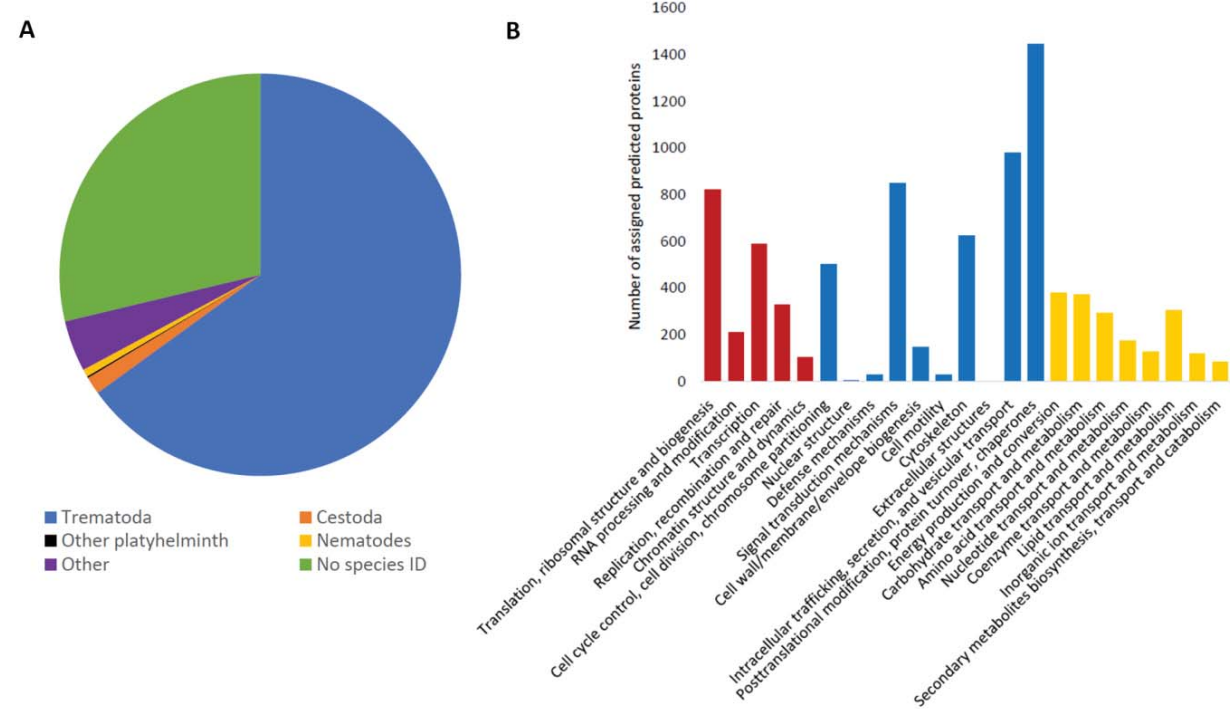


Figure 3.

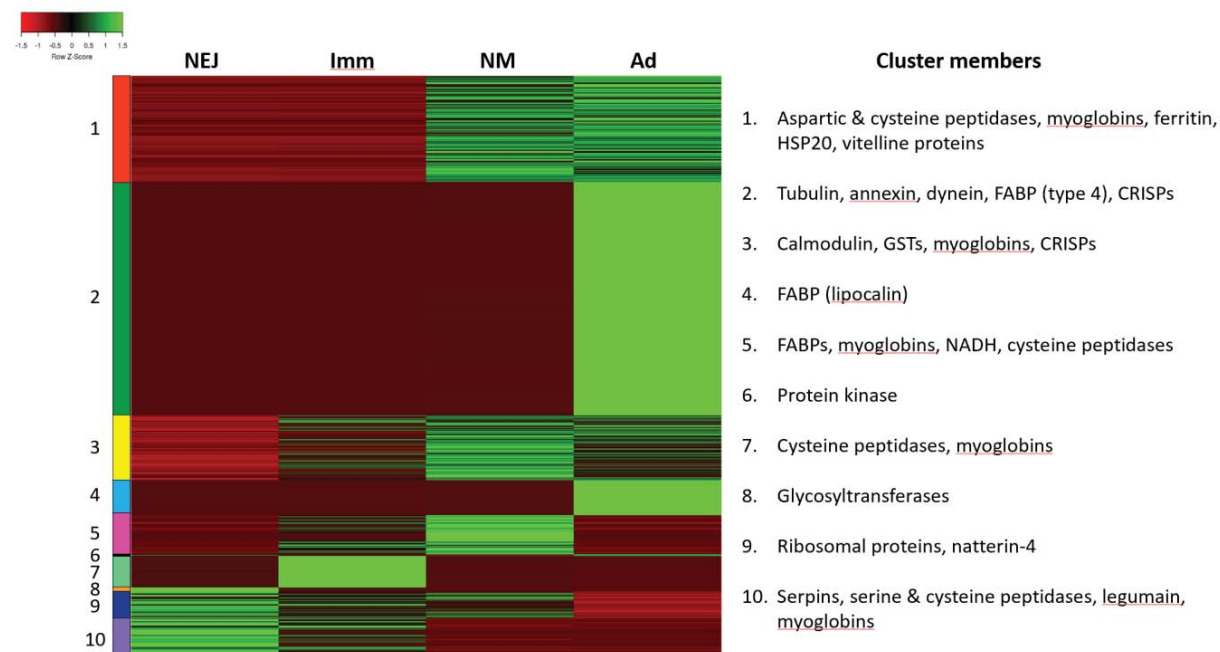




Figure 4.

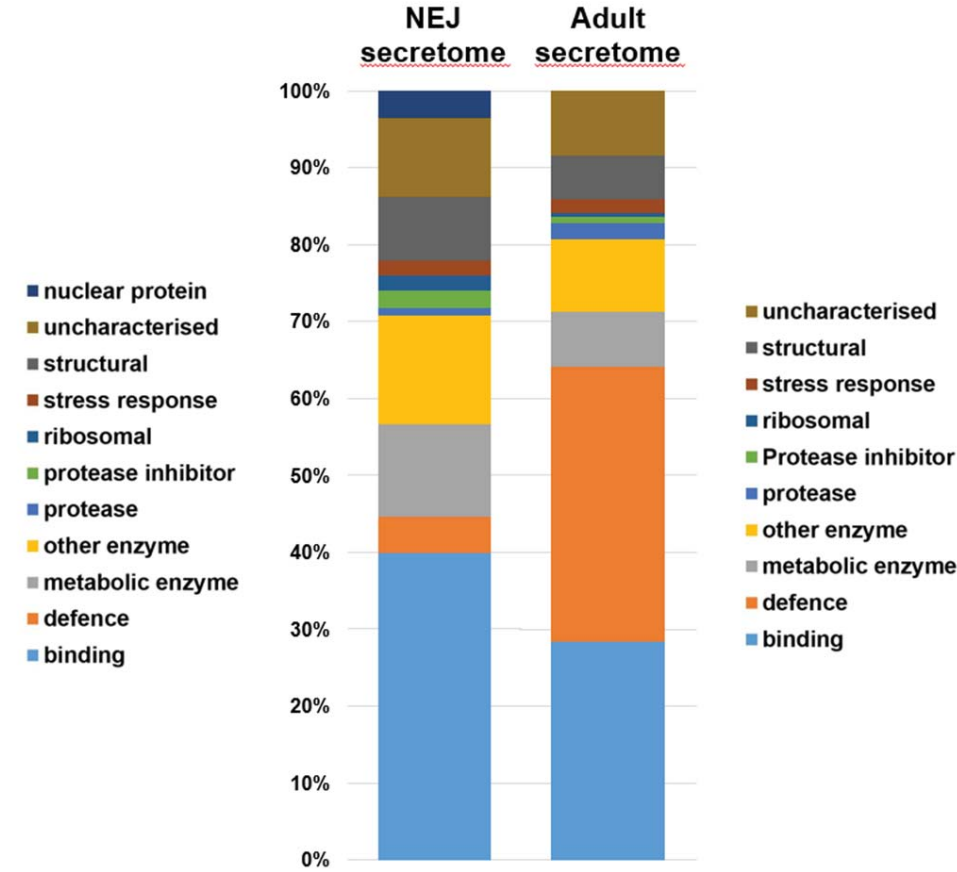


Figure 5.

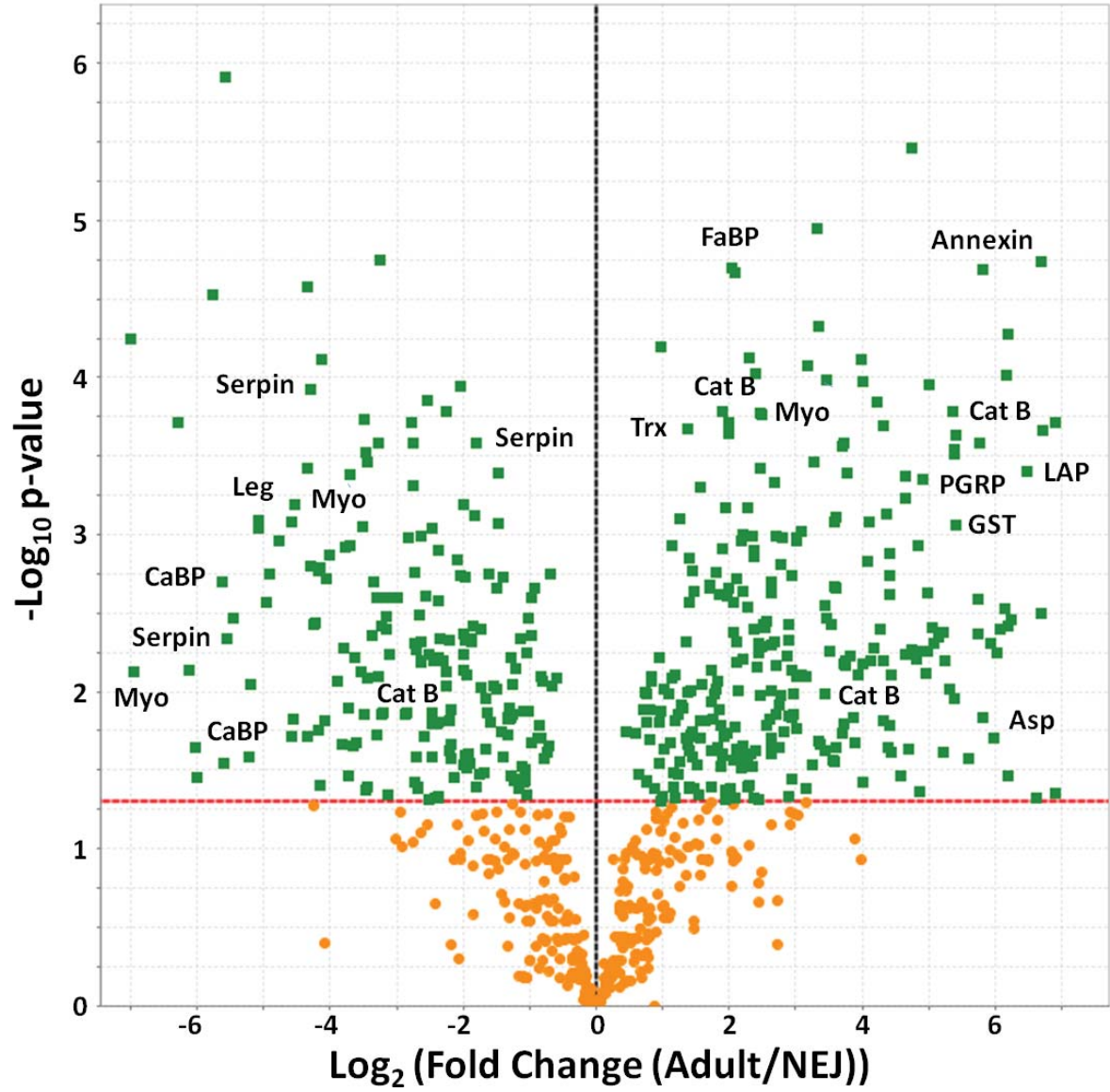


Figure 6.

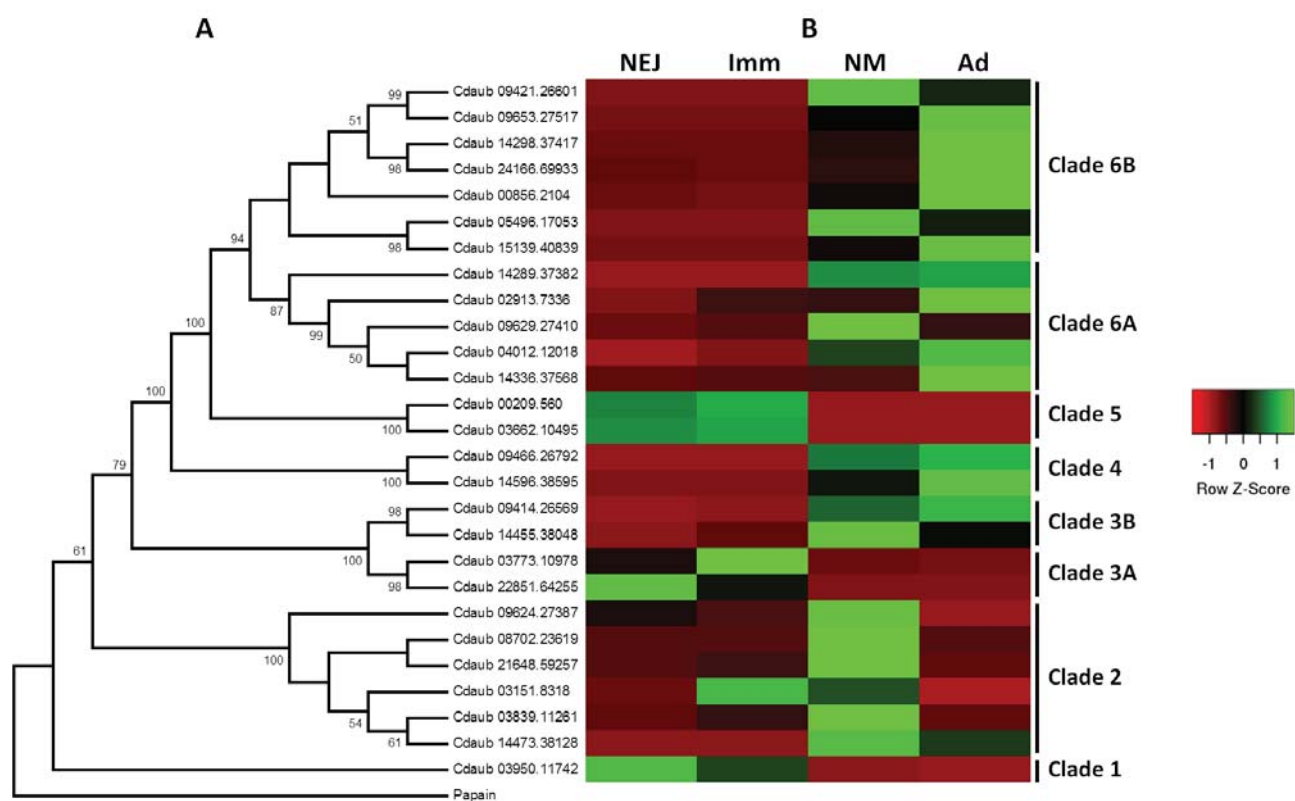


Figure 7.

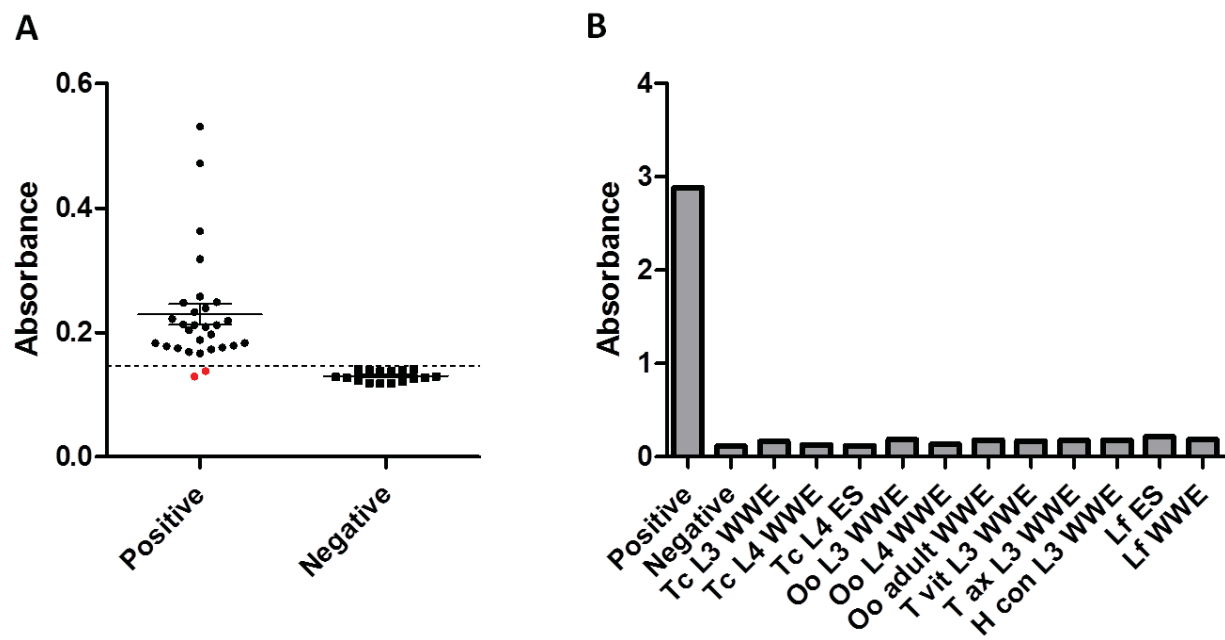


Figure 8.

